Analysis, Event and Group Prediction of Terrorist Networks using Computing Techniques

By

Wasi Haider Butt
2009-NUST-Tfr PhD-CSE-58

Submitted to Department of Computer Engineering
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in
Computer Software Engineering

Thesis Supervisor
Dr. Shoab Ahmed Khan

College of Electrical and Mechanical Engineering
National University of Science and Technology
2014

Acknowledgements

First of all, I am thankful to ALMIGHTY ALLAH, Who blessed me with strength and courage to complete this humble effort. Nothing can be done without HIS permission, so I believe HE has granted me to finish this work.

I would like to express my gratitude to my very kind advisor Engr. Dr. Shoab A Khan, Head of Department of Computer Engineering; from the deep core of my heart for all he has done for me to complete this work. Without his kind advice, encouragement, guidance and support, it was impossible for me to carry out this task.

I thankfully acknowledge, my co supervisor Engr. Dr. Farooque Azam whose kind guidance and constant support helped me a lot achieve this target. I am also thankful to members of my guidance and examination committee members from the core of my heart including Engr. Dr Usman Qamar for his endless effort and support for me especially in understanding data mining concepts, Dr Aasia Khanum for helping me a lot to understand various artificial intelligence concepts, both for college of EME. I am also thankful to Dr Sharifullah Khan from SEECS for being my external member of guidance and examination committee and for his continuous help and support especially in database concepts.

I would like to pay my very especial thank to Engr. Dr. Muhammad Usman Akram, who's intellectual and technical help and support always kept me on track. He always pushed me up whenever I felt disappointment.

I also am deeply thankful to my ex commandant Maj Gen (R) Muhammad Shahid HI(M), who motivated me from the start of my degree till the end. His query to me regarding my PhD progress always encouraged and motivated me.

I would also like to thank Engr. Dr. Muhammad Younas Javed, Dean, College of EME, a very kind, helpful and supportive teacher of mine; for his kind guidance and support. I greatly appreciate all my teachers, friends and colleagues especially Engr. Dr. Rashid Ahmed, Engr. Dr. Ali Hassan, Engr. Dr Saad Rehman, Dr Nazar Abbas, Engr. Dr. Farhan, Engr. Dr. Arslan, Mr. Jahanzeb ,Mr. Abdul Wahab Muzaffar. And Mr Sajid Gul Khawaja

I have no words to thank and acknowledge my papa and mom, whose prayers uplifted me always and made me able to do and complete such task. Especially to my papa whose constant motivation and support was the greatest support and motivation factor for me. I am thankful to my sisters and other family members as well for their prayers and continuous encouragement.

Abstract

A sharp rise in terrorist activities has motivated many researchers to device techniques of identification, analysis, detection, and prediction of terrorist activities using computing techniques. Terrorists secretly structure themselves in organizations to be more effective. Members in these organizations have to interact and communicate with others in order to plan and carry out haronious acts of terrorism. The pattern of these interactions and communication can reveal the secret structure of these organizations and can also help in predicting their plans of undertaking these activities. The current research proposes to employ techniques for the analysis of social networks to evaluate their applicability on these secretly formed and structured networks. As it can be viewed from the definition of social network which is "A social network is defined as a social collection made up of social actors like persons or organizations and a compound set of links between these actors". This definition entices us to view terrorist networks as social networks to apply social network analysis to extract their inner structure in form of useful knowledge. The inner structure would reveal the importance of each actor in the network and can then be used for suggesting counter actions that can help in easy destabilization of such organizations preventing them to carry out terrorist incidents. Based on our investigation, we could formulate the fact that traditional social network analysis measures are not directly relevant. This is because of their desires of hiding their intentions and links. Keeping such considerations in mind, this thesis proposes a new measure "Relative Degree" for terrorist network analysis. This thesis, builds on to this novel measure and the techniques for the analysis of the network and presents a model to detect active status of the network using outlier detection techniques on the communication/interaction or work patterns of these networks. An active state of a terrorist group or network is defined as the state in which the group is either planning or is already ready for execution on a worked out plan. The other state is passive, where the group is dormant and not executing any activities. The thesis proposes a technique where a percentage of communication is classified as outlier. These cases of outliers contain the active state of the terrorist network if contained. The technique is validated on a privately held record of cyber-attacks on an ERP system. To make the novel tool comprehensive for use the thesis further proposes a hybrid classifier for key player detection, This novel classifier has been tested on various publicly available and a privately held dataset. The technique gives an average accuracy

of 91.98% on available datasets. The proposed technique out performed once compared with individual classifiers.

The performance of the newly designed classifier is found satisfactory and up to the mark.

This thesis also proposes a novel Terrorist Group Prediction model. The model uses data classification of globally available historical data relating to act of terrorism for predicting the responsible terrorist group in a new incident. The classification is performed based on majority vote. The different options for the voting are the outcome of an ensemble of classifiers. The developed model is applied and tested on Global Terrorism Database (GTD), a publicly available dataset containing data of terrorist incidents occurred since 1970 till 2013, constructed by university of Maryland. The performance is calculated based on 10 fold validation that uses 10% of the data for testing and 90% for training in ten different iterations. The model achieves 93% accuracy that is the best accuracy once compared with the accuracies of the individual classifiers in the ensemble. To the best of our knowledge no such classification is performed on the dataset.

Table of Contents

List of Figures

List of Tables

Chapter 1 Introduction

Like many other fields, Area of National Security has also attracted number of researchers especially from Information Communication Technologies due to a number of reasons. Firstly National Security has gained vital importance since last decade because of the exponentially increasing number of terrorist events across the globe. Secondly studies have revealed that all of the terrorist events were not possible without use of latest information communication technologies as such events may not be possible without a very heavy collaboration among terrorists which definitely needs information and communication mechanism. Because of this reason, there is a room for effective research in order to contribute in the area of anti terrorism.

In this thesis, different computing techniques from the areas of Social Network Analysis, Outlier Detection and Data Classification have been proposed to be used in various counter terrorism processes. This chapter consists of motivation of selection of this specific topic and basic introductory concepts about the above mentioned fields. List of contributions and organization of the thesis has been given at the end of this chapter.

1.1 Motivation:

In the recent era, every nation and every country has suffered badly from terrorism. Unfortunately we belong to a nation and a country which is among the list of badly affected regions.

Figure 1.1 shows that Pakistan has always been affected with terrorism with problems in the neighboring countries. As our country has been suffered extremely badly so this became the motivation of choosing this topic for PhD research in order to put in my small contribution in the subject field.

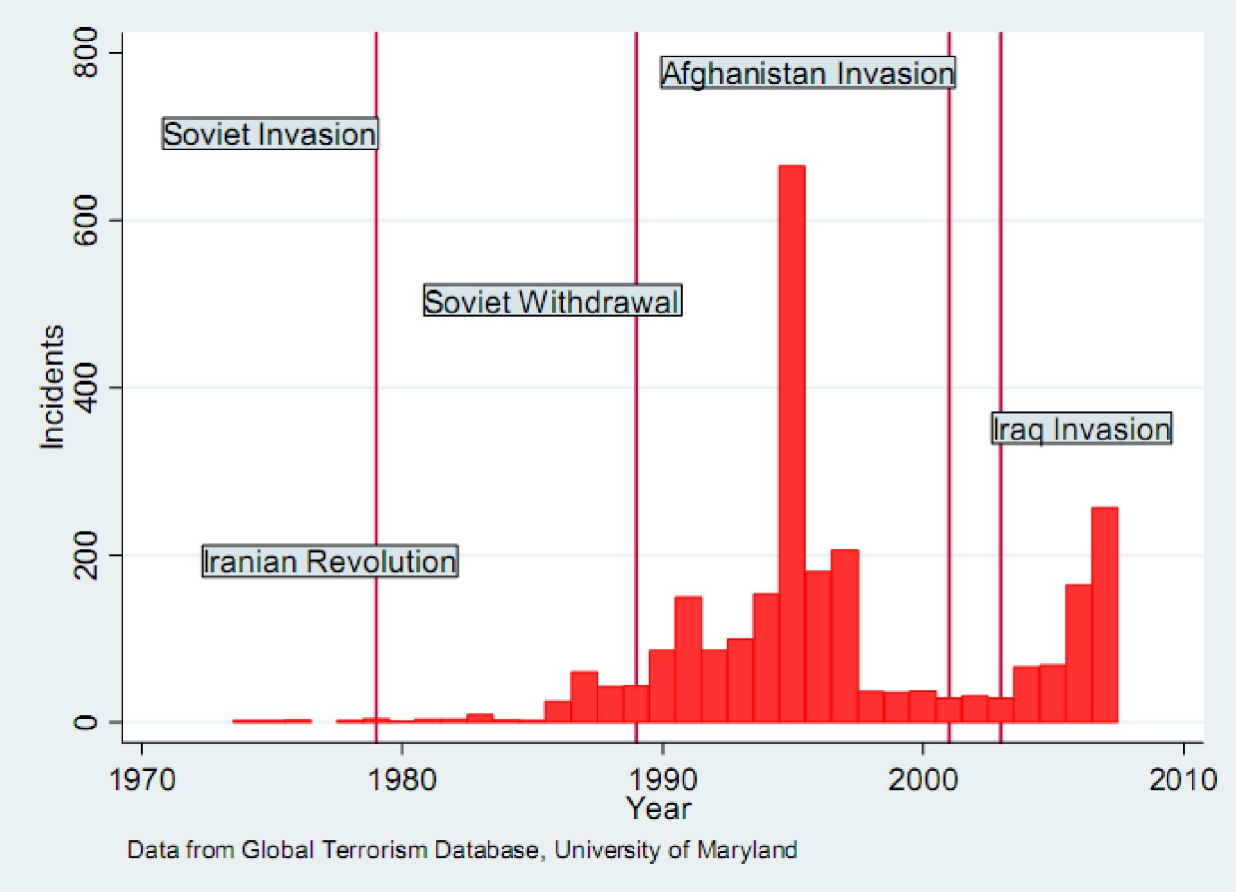

Figure 1.1 Number of terrorist incidents since 1973 to 2007 in Pakistan

1.2 Terror and Terrorism

The word terrorism is very hard to define; different sources provide different definitions of terrorism. A commonly used definition is "systematic use of violence and intimidation to achieve some goal".

A definition provided by UN General Assembly Resolution is:

"Criminal acts intended or calculated to provoke a state of terror in the general public, a group of persons or particular persons for political purposes are in any circumstance unjustifiable, whatever the considerations of a political, philosophical, ideological, racial, ethnic, religious or any other nature that may be invoked to justify them".

Terrorism affects social life of people including politics and management of events of any society. Terrorism in any form is harmful for any nation. It has affected the whole world. Terrorists working with different terrorist organizations have shed blood of a number of innocent people including women and children even. It destroyed number of communities, and made number of people homeless. Some worst very old and a new incident across the globe killing one hundred or more people are given below; number of casualties is given against each: [1]

- Bombing of Bolgard palace in Bessarabia (modern Moldova) 13 Dec 1921: 100 people killed
- Bombing of cathedral in Sophia, Bulgaria 16 Apr 1925: 160 people killed
- Mid-air bombing of Aeroflot airliner, Siberia 18 May 1973: 100 people killed
- Crash of hijacked Malaysian air liner near Malaysia 4 Dec 1977: 100 people killed
- Arson of theater in Abadan, Iran 20 Aug 1978: 477 people killed
- Hostage taking at Grand Mosque in Mecca, Saudi Arabia (includes 87 terrorists killed) 20 Nov-5 Dec 1979: 240 people killed
- Crash of Gulf Air flight following mid-air bombing over the UAE 23 Sep 1983: 112 people killed
- Truck bombings of U.S. Marine and French barracks, Beirut, Lebanon 23 Oct 1983: 301 people killed
- Armed attack on crowds in Anuradhapura, Sri Lanka 14 May 1985: 150 people killed
- Mid-air bombing of Air India flight off Ireland, and attempted bombing of second flight in Canada 23 Jun 1985: 331 people killed
- multiple bombings in Kano, Nigeria 20 Jan 2012: 178 people killed

Terrorists behind such deadly events often work in terrorist groups which we refer to as terrorist networks and are associated with one or more terrorist organizations. These terrorist

organizations have their own interests and targets which are normally hidden. These organizations are global. They seem to work across the globe rather than a specific isolated region. Table 1.1 shows fatalities in terrorist violence in Pakistan.

Year	**Civilians**	**Security Force Personnel**
2003	140	24
2004	435	184
2005	430	81
2006	608	325
2007	1522	597
2008	2155	654
2009	2324	991
2010	1796	469
2011	2738	765
2012	3007	732
2013	1985	427
Total*	**17140**	**5249**

Table 1.1 Causalities in Pakistan due to terrorism 2003-2013

Some well known organizations that have been included in the list of terrorist organization by US state department include: Abu Nidal Organization (ANO), Basque Fatherland and Liberty (ETA), Liberation Tigers of Tamil Eelam (LTTE), al-Qaida in the Islamic Maghreb (AQIM) and so on [2].

1.3 Social Networks:

A social network is defined as a social collection made up of social actors like persons or organizations and a compound set of links between these actors. The social network view provides a clear way of analyzing the structure of whole social entities [3]. Social Network Analysis is a mathematical method for 'connecting the dots'. SNA allows us to map and measure complex, and sometimes covert, human groups and organizations [4]. SNA has been applied in a number of applications in order to explore several interesting features of different sort of social networks especially with the advancement in information and communication technology and availability of social networks in electronic forms.

1.3.1 Terrorist Networks as Social Networks

Terrorist networks are the networks of different terrorists who work in organized structure in order to carry out different terrorist activities across the globe. Two issues are very important to discuss. The first is why terrorist networks should be counted in social networks at all and the second is that is a terrorist network a pure social network, i.e. can we apply typical SNA measures on such networks or are there any differences or specialties of such networks.

First answer to first issue is discussed i.e. why to include terrorist networks in the category of social networks. After a number of terrible incidents occurring across the globe, terrorist group's analysis attracted attention of number of law enforcement and other relevant agencies and also researchers especially after the disastrous event of 9-11. Different terrorist organizations' names appeared in print and electronic media. Analyzing all the information present in all available sources, one can safely conclude that terrorist organizations work as typical organizations with some specific characteristics and objectives. Members of terrorist organizations can easily viewed as nodes while there interactions, that can be of any type like communication, kinship, fellowship, attending training camps together and many more can be viewed as the links/ties/edges in a social network. Definitely in order to achieve the hidden objectives, members of terrorist organizations have to interact using different means. While considering second issue, i.e. are terrorist networks typical social networks, if we take a look of available literature, we will come to know that terrorist networks are not typical social networks although they can be treated as social networks but they have some specific characteristics which are

totally different from typical social networks. The behavior of covert networks is often not like normal social networks [5]. On contrary to normal social networks, here in covert networks the strong links remain mostly inactive and therefore hidden and only get activated when required [6]. In a normal social network, the most active nodes in terms of existing centrality measures are the key actors but it is not so in a covert network because of secrecy.

1.3.2 The Key Player Problem

An important area in SNA is the Key Player Detection. Key player is defined as the most important node in a social network. Importance criteria can be different depending upon the nature of that network. In social networks where individuals are not interested in hiding their relative importance in the network, finding them is comparatively simpler than in networks where individuals are intended to hide their identities, their roles and their importance for the stability of the network. Terrorist network are of the same nature where individuals are linked together in order to achieve some secret objective. All members are too vigilant from being caught. The problem is how we can find the most important actors from the network whose elimination can stop the whole network in achieving objectives.

1.3.3 Centrality

Centrality is a key theory in the study of social networks in order to study organizational and team behavior. Central individuals control the flow of information and decision making within a network. However, the connection between mathematical measures of centrality and the real world phenomenon of centrality is somewhat unclear. [7]

Nodes having higher values of centrality represents actors with the maximum structural importance in networks, and these actors would be expected to have a key role in simulated and real-world behavior. This applies to networks of many different kinds [3].

1.4 Outlier Detection

Outlier detection deals with detection of patterns from data which do not match to expected normal behavior. These anomalous patterns are often known as outliers, anomalies, discordant observations etc in different application domains. Outlier detection is a well researched area having an immense use in a wide range of applications like fraud detection, insurance, intrusion

detection in cyber security, fault detection in security critical systems, military surveillance for enemy activities and so on.

1.5 Data Classification and Prediction

Data classification means to classify some data instances into predefined classes according to the values of their features or attributes. With reference to classification, classifiers are the functions which fragment a set into two classes. The relation between classification and prediction is because of the fact that to assign proper class to a data instance is actually based on prediction. Prediction is based on training a part of the dataset in which class labels of data instances are already known. The classifier is trained using that known data and then prediction is made for classification of unseen elements on the basis of that training. Because of inclusion of the training step, classification belongs to supervised learning. This dissertation contains a complete chapter on classification and prediction, different classifiers and the standard measures used to evaluate the performance of different classifier.

1.6 Thesis Contributions

The research contributions made in this thesis are listed as under:

- A new social network analysis measure "**Relative Degree**" to detect group leaders in terrorist networks
- Application of Outlier Detection for event prediction suspicious networks
- Construction of real dataset and used for experimentation of outlier detection for event prediction
- A novel **hybrid classifier** based key player detection ensemble model
- A novel **hybrid classifier** for prediction of responsible terrorist group in a terrorist incident.

1.7 Dissertation Organization

The rest of this thesis is structured as follows:

- **Chapter 2: Relative Degree**. An overview of social network analysis, current measures used for key player detection and application of social network analysis in the field of

terrorism, Proposed Relative Degree, Experiments and Results are presented in this chapter.

- **Chapter 3: Outlier Detection and Event Prediction.** This chapter consists of basic concepts about outlier detection and its applications and techniques which exist in literature. Also the proposed application of outlier detection for event prediction is presented. Chapter also covers experimentation and results
- **Chapter 4: Classification and Prediction.** This chapter discusses concepts of data classification and prediction. The different existing methods of classification are also discussed. A hybrid classifier for key player detection and a hybrid classifier for terrorist group prediction are also presented. The chapter also contains experiments and results.
- **Chapter 5: Conclusions and Future Works:** This chapter concludes the work done in this dissertation. Directions about how to continue this work in future are also given.

Chapter 2 Social Network Analysis and Relative Degree

This chapter covers basic concepts and work done in the area of social network analysis (SNA), especially on terrorist networks. First section covers SNA in general and the key player detection in specific. Analysis of terrorist networks using social network analysis is covered in the second section. A new social network analysis measure, Relative Degree is proposed in the third section.

2.1 Social Networks:

As defined in the previous chapter, A social network is defined as a social collection made up of social actors like persons or organizations and a compound set of links between these actors. This section covers social networks and their analysis in detail.

Typically social networks are defined as graphical representation of social relationship between individuals or organizations where each node (also known as actor or vertex) represents an individual or group of individuals while an edge connecting two nodes (also known as tie) represents relationship between the objects represented by the two nodes. Usage of graph to represent social data has a number of benefits. It enables analysts to completely analyze the social network visually and to apply desired operation clearly and conveniently. Also all graph theoretic concepts can be easily applied in order to analyze the social network mathematically.

Generally social networks can used to symbolize, recognize, and calculate any type of correlations between any kind of entities, such as words, web pages, people, organizations, animals, cells, computers, and other information or knowledge processing entities [8]. Thus, social networks have broad and successful applications in sociology, epidemiology, biology, criminology, and economics [9].

An example school fellowship relationship is shown in a matrix form in Table 2.1. In both rows and column, headers are the individuals i.e. the nodes or the actors while at junction of two nodes i.e. in the cells of table is a binary values representing either fellowship relation exists or not. In the diagonal, all null values represent that we are not interested to model either an individual has a fellowship relation with itself or not.

	Ahmed	Ali	Raza	Abbas	Shahid	Michael	Susane	Asif	Wahab
Ahmed	--	0	1	1	0	0	0	0	0
Ali	0	--	0	0	0	0	0	0	0
Raza	1	1	--	1	0	0	0	0	0
Abbas	1	0	1	--	0	1	1	1	0
Shahid	0	0	0	0	--	1	0	0	0
Michael	0	0	0	1	1	--	1	1	0
Susane	0	0	0	1	0	1	--	1	0
Asif	0	0	0	1	0	1	1	--	0
Wahab	0	0	0	0	0	0	0	0	--

Table 2.1 The social data about school fellowship relation between individuals

Table 2.1shows the school fellowship relation. A binary 1 between two nodes which are shown in row and column headers represents that those two individuals have attended school together while a zero shows no relationship. A matrix can be used to represent a complete graph i.e. a complete social network as can be viewed clearly from the above table.

The information from this table can be converted into graph for visual representation of the network at anytime. Figure 2.1 represents the graphical view of the network.

Social network gathering can be mainly divided into two broad categories. One is elicitation and the other is registration [9]. In elicitation, relationship information is gathered through questionnaire/survey etc while in registration relationships information is extracted from registered information like membership lists, email records, SMS logs, telephone call logs etc

Traditionally in social network analysis research, elicitation methods were used. In elicitation methods questions about relationships are asked and population is required to answer the questions generated. It seems quite simple to develop some questions and then simply get them answered by the individuals but data gathered through such methods may be quite inaccurate and subjective [10].

Complete data cannot be gathered using just questionnaires or surveys. Also data can get affected by biasness of target audience. Secondly every individual will have its own perception

and will respond according to his/her understanding. So the definition of relationship will be fuzzy.

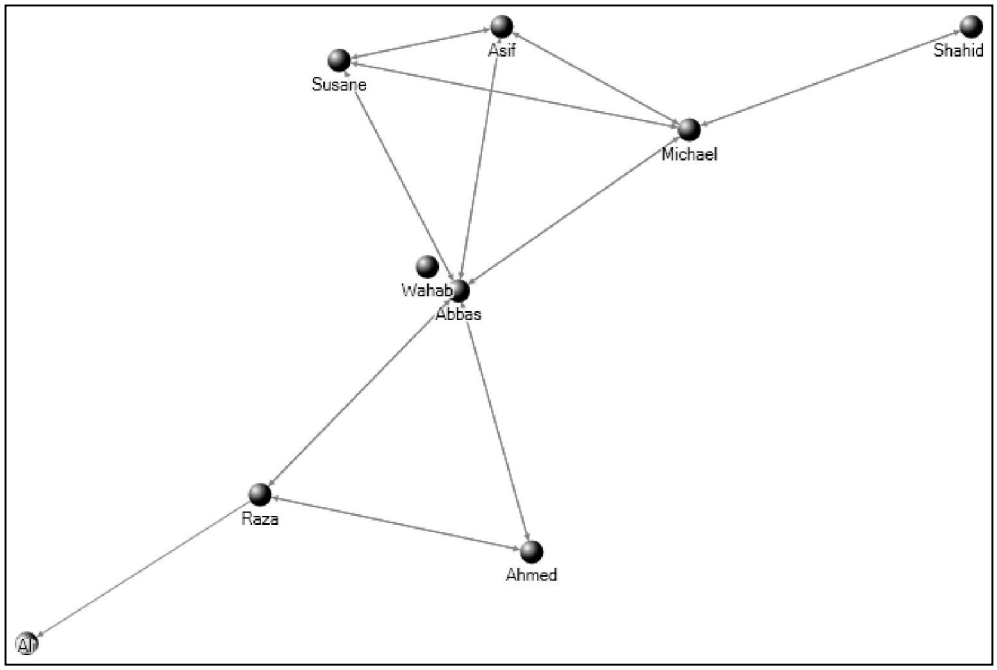

Figure 2.1 A School fellowship network

Other than these problems, these methods will also imply high labor costs. Especially when the target network consists of number of nodes, it becomes almost impossible for researchers to collect data using these methods. Another issue related to these methods is the nature of target network, e.g. the network under study may not be of such type in which information can be gathered using such methods. Consider, for example, if network under study is a smugglers' network, it will become practically impossible to collect data using such methods.

With advancements in the field of information technology, mostly automated data collections are now used in almost all the fields. It has become possible because of the storage of relationship information in electronic form. For instance at email servers, data associated with an email is stored electronically which consists of sender and receivers. This data associated with a typical email show a communication relationship between the individuals associated i.e. the sender and the receiver of that specific email.

Along with other advantages such methods of data gathering in social network analysis have also an edge of having more objective definitions of relationship over the traditional techniques. For

example if an email has been sent from a person to the other, the relationship is clear. Criteria can also be set for instance if a sender sends more than ten email to the same person, it may be defined they have a trust relationship between them, so criteria definition is easy to set up and implement as compared to traditional methods.

2.2 Social Network Analysis (SNA)

How to extract, manipulate and manage structural information present in social network comes under the field of Social Network Analysis. Different researchers have defined SNA different but some relevant manners. For example according to Krebs, SNA is the mapping and measuring of relationships and flow between people [8]. Similarly according to Freeman, SNA are the techniques focusing on uncovered patterns of people's interactions [11]. Scott defined SNA as a set of methods for the investigation of relational aspects of social structures [12].

All of the above mentioned definitions accentuate that SNA deals with the study of structural information embedded in the interaction/relationships between the nodes of a social network structure. SNA primarily deals with the study of relationships between nodes instead of nodes themselves. We can also say in terms of social network structures that SNA primarily deals with the analysis of edges connecting the nodes other than focusing on just nodes or their attributes. This doesn't mean that attributes of nodes are not of any importance. Many times these node attributes help analysts in having understanding of social behaviors and analysis of specific social phenomena. So nodes and their attributes are equally important but the primary focus of SNA is the analysis of structural properties and interaction patterns between nodes.

SNA analyzes the structural characteristics of individuals or groups of individuals in a network. The analysis consists of how individuals are linked with each other in a network, how an individual can affect links between other individuals and how different groups of individuals are linked together. Broadly SNA analysts are also interesting in finding characteristics like the connectivity of the network, decomposition of network into subgroups on the basis of relationship characteristics etc.

Traditionally if we are interested in finding how a group operates, we may go to the organogram of the organization and will find the people sitting at the top. The people who have higher position in the hierarchy are empowered to take decision so they will be considered effective and

influential but in evolving networked organization, organogram is no longer an adequate guide to have an understanding of operating of a group. On the other hand SNA statistics will provide us with true picture of who is most active and most influential in the group who can disseminate information amore quickly and accurately no matter he might be sitting at the lowest position in the hierarchy of the organization.

Normally SNA consists of three mail elements. The first is a group which may consist of individuals who are dedicated to a specific task e.g. people working in finance department, or it may be a collection of individuals such as a community. Individuals having some common attributes of our interest will be contained in one group. The members of a group are the individuals or actors. Second element is the relationships among individuals which is also the main focus of analysis. The nature of relationship under analysis depends upon the reason of conducting the analysis of a group or network. Interactions are also known as links or ties between individuals. Basically a social network is the pattern of interactions between individuals so importance of interactions in SNA is clear from this statement. The third and last element is the attributes which are not the primary focus in SNA however the attributes that can help conclude whether there are logical factors that affect relationships between individuals are part of the analysis. So only the attributes that are supposed to influence the interactions between individuals are included in SNA.

Because of increasing interest in the field of SNA and also with advancement in Information Technology, researchers and other software corporations have put in huge effort in developing computer software that have made SNA easier, convenient and efficient. Many open source and proprietary SNA tools are available in market now days. The tools help visualize networks graphically and also to perform analysis providing ease of applying several SNA operations on any network.

Some analysts believe that human eyes are the best analystic tool, so some SNA tools have been developed keeping this thing in mind and they provide network visualization in a very convenient manner e.g. NetDraw, NetMiner, and Pajek. On the other hand, others rely on text reports generated by tools consisting on SNA measurements and analysis. Examples of such tools are UCINET, Agna, and MultiNet. A hybrid approach can give best results. Some tools also provide statistical analysis of networks like StOCNET. Software support for SNA is still a

developing field. Still the tools require improvements and still a lot of efforts are made in the field. A very critical problem is of complexity that is often seen while analyzing large networks with the help of these tools. Different tools impose different limitations over the number of nodes and relations that can be analyzed. However new versions arise day by day overcoming this problem of complexity.

2.3 Application of Social Network Analysis

As discussed earlier, a social network is a social structure between individuals or organizations indicating the way in which individuals are connected. Different networks like Email traffic, disease transmission and criminal networks can be modeled as social networks. The analysis of any such network which can be modeled as a social network is social network analysis (SNA) as discussed earlier in the previous chapter. SNA deals with the mapping and measuring of relationships between individuals of a social network. Figure 2.2 shows network of AIDS.

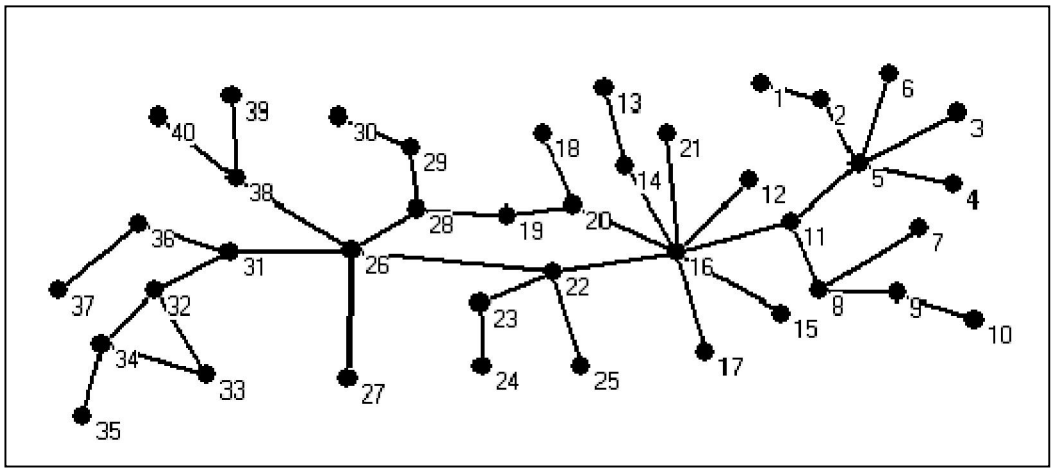

Figure 2.2 The sexy relation network of the AIDS [32]

SNA provides both visual and mathematical analysis of relationships between nodes which can be generally individuals or organizations. Another important term is organization network analysis which is a form of SNA in which the social network under study is an organization. An important study in SNA is the study of sub groups also known as cliques in a social network. Normally the size of cliques in terms of nodes and their relationships and interconnections between cliques is studied because this study can tell a lot about the behavior of the network as whole.

SNA is normally applied to know things like how fast or slow a resource or information can flow through the network? How conflict may be created and how will they affect the network? What is the magnitude of overlapping of different cliques? And so on. Data about social networks consists of a number of elements. Social networks data can be viewed as a social relational system characterized by a set of actors and their social ties [3]. A very important area in social network analysis is the detection of most important nodes from the whole network. This detection can be helpful in many applications e.g. detection of most influential members of a group can help covering a whole group, detection of such most involved nodes is a sexy network can help control sexually transmitted diseases, detection of such nodes in terrorist networks can help destabilization of such groups and so on. Key player problem is discussed in detail in the next section.

2.4 Key Player Detection

A very important area in social network analysis is detection of key players from a social network. Definition of key player can be different for different networks and from different points of view. Let us define key player concluding all the available literature as, "Key players in a social network are the nodes which are relatively more important than other nodes in the network". Key players are members of a subset of a social network which consists of some important nodes. Importance criteria can be different depending upon the network and mainly on the type of analysis.

A very well known name Stephen P. Borgatti defined the key player problem consisting of two separate problems. Borgatti follows the problem specified approach to the key player problem introduced by Friedkin [13]. According to him the definition of key player depends on the requirement. We also accept and follow his approach in defining the key player.

Borgatti [14] defined two totally different aspects of the key player problem. The first key player problem is defined in terms of the degree of dependence of the cohesiveness of the network on its key players. This has been referred as Negative Key Player Problem abbreviated KPP-Neg. In this Key player problem, importance is measured in breach i.e. the amount of reduction in cohesiveness of the network that will occur if the selected nodes are removed from the network.

The second key player problem is defined in terms of the degree to which key players are connected to and embedded in the network around them. This is referred as positive key player problem abbreviated KPP-Pos.

The two discussed approaches can be defined generally as i) In a social network, detecting a set of *k* nodes which if eliminated would maximally disorder communication among the remaining nodes and ii) In a network, find a subset *k* that is maximally connected to all other nodes [14].

One very important measure used traditionally in detecting the most important nodes in SNA is measuring the centrality of all nodes present in the network and then setting a threshold for the score of centrality measure used against each individual, selecting nodes crossing that threshold. Centrality is defined in the next section in detail.

2.5 The Problem of Centrality

The concept of centrality as applied to human interaction was introduced by Bavelas in 1948. His specific concern was communication in small groups. He hypothesized a relationship between structural centrality and influence in group processes [15]. The first research application of centrality was made under Bavelas's direction at MIT in late 40s. The first study was conducted by Harold Leavitt (1949) and Sidney Smith (1950). It has been seen that the network centrality philosophy is very powerful for solving problems involving large complex networks [16].

Centrality is the level to which an individual is in the center of a network. Central people are more influential in their network. In business organizations such central individuals are likely to get better performance reviews, and generally are more satisfied with their jobs than individuals who are less central. Although a lot of research has been done in defining measures of centrality but the most well known measures are degree, betweenness and closeness centrality. There are other measures also which different researchers have proposed, some of them are specific for some domain while others are generic like the basic measures discussed above.

Centrality is measured in order to find the most important nodes in the network. The prime supposition behind this is that the more central position a node has got in the network, the more important its role is. This will be discussed in detail in the coming sections. Other than being central in the network, importance criteria may be different depending on the study, the network under study and the focus of analysis. The available data source can be also an important factor.

Some researchers have also proposed other ways of finding the most important nodes. So in this chapter first the centrality measures are discussed then other ways of finding the most important nodes are also presented.

2.5.1 Degree Centrality

The Degree Centrality DC is simplest centrality measure. It uses the number of direct contacts of a node as a pointer of the quality of interconnectedness [17]. Using the adjacency matrix A = (a_{ij}) it can be formalized as shown in Equation 2:1

$$\sigma(v) = \sum_{i=1}^{n} a_{iv} \qquad \textbf{Equation 2:1}$$

Degree Centrality of a node in a network is the simplest structural property simply counting the number of neighbors. For undirected networks where edges only have magnitude without any direction, Degree Centrality is simply the measure of neighbors of each. While for directed networks when edges also have a direction, two different type of degree can be defined. The first is in-degree which is count of edges pointing towards a node whose in degree is being calculated and the second is out degree which is the count of edges point towards other nodes from the node whose centrality is under calculation.

In degree centrality the number of direct contacts of a node is considered to be the indicators of the quality of a network member's interconnectedness [17]. The node having more connections has a high value of degree centrality. So a high value indicates more central position, hence more importance in the network.

A high value of degree gives a high rank to a node in the network while taking degree centrality as criteria for detecting the most central nodes.

The application of degree centrality is in the areas in which the purpose of analysis is to detect where the action in the network occurs and where the criteria for importance is the one hop connectivity of nodes. For example in case of a social networking website, a user with highest number of friends will be considered most central and influential due to highest number of connection hence highest degree. Similarly in case of analysis of a society for a particular sexually transmitted diseases, the person with highest relations will be the most vulnerable, so degree centrality will be applied to find out most vulnerable person. Some analysts have also

applied degree centrality in terrorist networks and they believe that the person who has highest number of connections in the group actually leads the group and his removal can destablilize that terrorist group.

The major disadvantage of degree centrality is that indirect contacts are not considered at all. So if an indirect neighbor of a node with highest centrality is removed, which may affect the network, there will be no affect on the value of degree centrality of that node.

2.5.2 Weighted Degree Centrality

The extension of degree centrality for weighted networks is known as node strength [18]. So weighted degree centrality is simply sum of weights of all ties adjacent to the node under consideration as shown in **Error! Reference source not found.**.

$$\sigma(v) = Strength(v) = \sum_{i=1}^{n} w_{iv} \qquad \textbf{Equation 2:2}$$

Where w is the weighted adjacency matrix. The value of w_{iv} is greater than zero if node 'i' is connected with node v. The value of w_{iv} represents the weight of the tie. Weighted degree gets equal to degree centrality if the network is binary i.e. if tie link has weight equal to 1. In weighted networks, the outcome of these two measures is different. As node strength takes into consideration the sum of weights of all links connected while degree considers only the number of links, so measurement gets different. Consider a node in the network which has 100 links, each of weight of 1, the weighted degree of that node will be equal to on hundred, on the other hand if in the same network, there is a node with only one link with a weight of 101, now weighted degree of this node with only one link becomes greater that the first node with one hundred links which is against the basic Degree Centrality.

As degree and strength (Sum of weights) can both indicate the rank of involvement of a node in the adjacent network, it is important to integrate both these measures when studying the centrality of a node [19]. To combine the degree and strength for weighted networks, a tuning parameter α has been used in [19] which decide the relative significance of the number of links compares to link weights and the idea is presented as a new degree centrality measure in the same work.

2.5.3 Closeness Centrality

As discussed above degree centrality only considers the directly connected nodes and not the global information, so here gap for new measure is created. Often there comes a requirement to score nodes based upon more global structural information. To score nodes based on global information, Closeness or Distance centrality is used [20]. This measure is based on how close a node is to other nodes in a network. For nodes which are at a shorter distance to other nodes in a network, closeness centrality has a higher value. Formally distance centrality can be represented as [21], as shown in Equation 2:3.

$$\sigma(v) = \frac{1}{\sum_{i=1}^{n} d(v,i)} \qquad \textbf{Equation 2:3}$$

Closeness Centrality determines how rapidly an actor in a social network access information through other actors of the network. A relatively higher closeness value of a node depicts that the node has a shorter path to other nodes in the network and so can reach them for information propagation briskly. Any individual at such position in a network i.e. having a higher value of closeness will be having more visibility to what is going on the whole network and hence can also control the whole network more effectively.

Normally Closeness Centrality is calculated using geodesic distances which are the shortest paths in the network. So in terms of distance, the nodes with high closeness centrality are the nodes which have shortest paths to the other nodes in the network.

The method of measuring closeness centrality is by taking sum of the distances of the node under consideration to all of the other nodes in the network and then by taking inverse of the average of that calculated sum as can also be seen from the mathematical formula for calculation of closeness centrality given above.

2.5.4 Betweenness Centrality

Using Betweenness Centrality, a node in a network is considered more central if it is located on many shortest paths between pairs of other nodes. The basic idea behind this centrality measure is that the interaction between two indirectly connected nodes i and j depends on the nodes between i and j. The formulation of Betweenness Centrality is given by Freeman as [15] as shown in **Equation 2:4**.

$$\sigma(v) = \sum_{i=0, i\neq v}^{n} \sum_{j=1, j<i, j=v}^{n} g_{ij}(v)/g_{ij} \qquad \textbf{Equation 2:4}$$

Higher value of betweenness centrality for a node indicates that the node occurs most often on the shortest paths between other nodes. So the nodes which are more probable to be come between two nodes on their shortest path have higher betweenness centrality. Nodes with such behavior i.e. having high betweenness centrality can also be described as gateways. Such nodes have a control over the data flow between different nodes or groups of nodes in the network because of the bridging nature. Such nodes can be vulnerable for the destabilization of whole network being single point of failure and being communication bridges. So if communication of whole network is to be disrupted, nodes with high betweenness can be targeted and eliminated.

The idea of this centrality was first introduced by Bavelas in 1948 [22]. He proposed that when a specific individual in a group is strategically located on shortest communication path connecting pairs of others that individual has a central position in the network and is more important than others. Other individuals of the network are assumed to be responsive to individual in such central position who has a hold over information flow and can hence manipulate the whole group by withholding information or coloring or destructing it in transmission.

Shimbel in 1953 expressed the same concept about the centrality measure but put it in quite different terms [22]. He described that if two nodes have to communicate and they must have to use a third intermediate node that comes on the path between them, then the middle node has a certain responsibility node the communicating nodes. If for the same intermediate node, all minimum paths that pass through that are counted; a measure of stress will be the result which that intermediate node must undergo during the network activities. So such a measure in numbers for every node of the network gives us a good idea of stress conditions throughout the network. So a node which has higher value of such measure has higher stress over it but on the other hand has a control. Shaw has termed such intermediate individual as relayer for a specific path. So relayer has high stress on it but has also a control over the communicating nodes and can influence them by refusing to pass their information from itself

If network under study is of the nature that working mainly depends on the flow of information then betweenness centrality can be chosen as the criteria to find most important nodes. Betweenness centrality has become a popular approach to deal with complex networks for last

few years especially in fields of computer and social networks, biology, transport, scientific cooperation, and so forth.

2.5.5 Eigenvector Centrality

Eigenvector Centrality is based on the idea that a connection to a more interconnected node has more impact to the own centrality to a greater extent than a connection to a less well interconnected node. For a node v, the EC is therefore given in **Equation 2:5** [23]:

$$\sigma(v) = 1/\lambda_{max}(A) \sum_{j=1}^{n} a_{jx}.x_j \qquad \textbf{Equation 2:5}$$

With $x=(x_1,\ldots,x_n)^T$ referring to an eigenvector for the maximum eigen value $\lambda_{max}(A)$ of the adjacency matrix A.

In other words Eigenvoector can also be defined to be proportional to the sum of the centralities of the node's neighbors so that node can have higher importance either by being connected to a lot of others or by being connected to others that themselves are highly important

Eigen vector centrality is simply the value of all nodes calculated from Eigen values which are a linear algebra concept.

2.5.6 Valued Centrality

Valued centrality was presented as an substitute to closeness centrality [24]. It was originally proposed for valued network considering tie strength but is also applicable to ordinary network [7]. It is defined like closeness centrality but unlike closeness, is the average of reciprocal rather than reciprocal itself. Mathematicaly it is given by **Equation 2:6**

$$\sigma(v) = \frac{1}{n-1}\left(\sum_{v\neq i} \frac{1}{d(v,i)}\right) = AVG_{v\neq i}\left(\frac{1}{d(v,i)}\right) \qquad \textbf{Equation 2:6}$$

Valued centrality is obtained by averaging the closeness values. The consideration upon which valued centrality is based is that closeness is the inverse of distances as mentioned in [24].

2.5.7 Jordan Centrality

Hage and Harry proposed the concept of Jordan Centrality based on the concept of centrality (Jordan Center) discovered by Camille Jordan in 19th century [25]. Only the largest distances are

taken into account in order to find the relative importance of each node in the network. Mathematically Jordan Centrality is given by **Equation 2:7**

$$\sigma(v) = \frac{1}{MAX_{i \neq v}\, d(v,i)} \qquad \textbf{Equation 2:7}$$

In contrary to typical analysis of social structures where purpose of analysis is to choose nodes for a facility on the basis of a specific criterion, in some cases while analyzing a social network, the purpose of analysis is to choose a node which minimizes the travel time between the node and all other nodes. The first problem is solved by finding the set of nodes whose total distance to all other nodes is minimum (Mathematically the median of the graph) as we did see in closeness centrality. On the other hand the second problem is solved by discovering the nodes whose maximum distance to any other node is least mentioned the center of the graph [24].

The eccentricity e(v) of a node v defined in [24] in a connected graph G is the maximum distance d(v,u) for all u. The diameter of a graph G is the highest eccentricity of a node i.e. the highest distance between two nodes of graph G. The radius r(G) is the smallest eccentricity of the nodes. A node is central if e(v)=r(G) and the center of G is consists of all such central nodes.

2.5.8 Flow Centrality

Another centrality measure "Flow Centrality" [26] proposes an alternative to betweenness, more suitable for valued networks. In betweenness centrality nodes are given positional advantage i.e. their importance increases with increase in the extent to which they lie in the shortest pathway between other pairs of nodes. So the basic idea is that how more nodes are between other nodes than other nodes. So this broker role is considered into power, more paths a node has through it, more control and power it has. If a node that has a broker role in communication of two other nodes refuses to let information pass through it, the sender and receiver must have to stop communication between them through that node, but they may exist some other paths through other nodes which may not necessarily be the shorter ones. So the flow centrality expands the idea of betweenness centrality. It is based on the assumption that nodes will use all the paths that connect them rather than relying completely on the shortest paths. So centrality is measured by the proportion of the entire flow between two nodes through all paths connecting them that occur on path of which a given node is part. For each node then the measure adds up how involved that

node is in all of the flows between all other pairs of nodes rather than computing just for shortest paths. The measurement number for flow centrality will grow with size and density of the network.

Mathematically flow centrality is given by **Equation 2:8**.

$$\sigma(v) = \frac{\sum_{i \neq j} \sum_{i<j \neq v} m_{ij}(v)}{\sum_{i \neq j} \sum_{i<j \neq v} m_{ij}} \qquad \textbf{Equation 2:8}$$

2.5.9 Trust centrality

Previous centrality measures are based on analyzing the influence or importance of a node depending on its position in the network. So they are only bound to the interactions with in the social network. The basic idea behind trust centrality is that if communication between nodes of a network are not bound to follow the edges, the network influence will be derived by trust. Trust behind trust centrality is totally personal. Confidence is required by one node to trust other that the information shared by a node to the other will not be misused by that node. So such type of trust is more personal. If a node A trusts B then A is personally affected if he is wrong in his trust (if his trust is broken e.g. his personal data is misused.

Trust centrality proposed in [27] depends on the privacy settings of each node in the given network. For all nodes in the network, relative trust values are computed and summed, trust value is decided by a trust function which depends on privacy settings of other nodes for this particular node. Implicitly every node shows on whom in the network it is trusting and on whom not by adjusting privacy settings. According to [27], trust based on privacy settings is computable. Trust centrality is given by **Equation 2:9**

$$\sigma(v) = \sum_{u \in V} (w_1(u) * \tau(v,u)) \qquad \textbf{Equation 2:9}$$

Where τ is a trust function showing for each pair (v,u) a measure for the trust shown by v for u. Trust centrality is the weighted sum of the trust shown by each node in the network. So according to Trust Centrality, a node is more central and important if more nodes rely on it and how much a node is trust worthy in a network can be quantified and computed as discussed above.

The more influential a node will be if value of others' trust on it is high and will be less influential if trust has a lower value according to trust centrality.

2.5.10 Dependence Centrality

Dependence Centrality of a node in a network is the measure of its dependence on any other node of the network [21]. Dependence Centrality is given by **Equation 2:10**.

$$\sigma(v) = \sum_{v \neq u, u \in G}^{n} \frac{m(v,w)}{N_p} + \Omega \qquad \textbf{Equation 2:10}$$

Where v is the root node which has dependence on node u, N_p is the number of shortest paths from node v to node w passing through node u, m(v,w) is the inverse of geodesic distances 1/d(v,w) from node v to node w. Value of Ω depends upon connectivity of graph, the value is 1 if graph is connected and 0 if disconnected.

The first part of the formula calculates that how many times a node v uses node u to communicate with node w of a network.

The idea has been applied on terrorist networks taking 9/11 hijackers network for experimentation in [28].

2.5.11 Using Tunable Centrality Parameters

It is a general conclusion that nodes with higher degree centrality also have higher betweenness centrality definitely because larger the degree of a node, larger the chances that many of the shortest paths will pass through this node but there are some exceptions. The importance of a node can not only be taken from its degree and betweenness but also from degree of its neighbors. A new way as a solution to this issue is presented in [29]. The new measure of node importance with tunable parameters is introduced in this research. Importance of a node v according to the proposed idea in [29] is defined in **Equation 2:11-**

$$\sigma(v) = \alpha d(v) + \beta d\tau(v) + \gamma B(v) \qquad \textbf{Equation 2:11}$$

Where α, β, γ are tunable parameters, and according to node role, α > β > γ usually. For simplicity α + β + γ have been set equal to 1. $d(v)$ is the degree of node v, B(v) is the value of

betweenness centrality for the same node and $d\tau(v)$ is defined as the total degree number of neighbors of the same node.

The proposed idea has been implemented and test on sexy networks and proposed formula has been compared with other centrality measures using different values for α, β and γ.

2.6 Other Ways of Finding Important Nodes

Other than fundamental centrality measures to find out the key players or important nodes in a given network, researchers have also introduced other means to solve the same problem. Some of those are derived from the above mentioned fundamental SNA measures and some are completely different approaches.

2.6.1 Using Graph Entropy

Discovering Important Nodes through Graph Entropy has been discussed by Jitesh Shetty and Jafar Adibi [30]. The graph entropy used is taken from Korner definition [31] when the graph is complete. To solve the problem of discovering important node in a network, the authors of the article have exploited an information theoretic model that according to them combines information theory with statistical techniques from the area of text mining and natural language processing.

How entropy models on graphs are relevant to study of information flow within an organization has been shown. Following is the algorithm proposed to use graph entropy for detecting important nodes:

So the nodes whose removal effect network the most are most important nodes according to this concept.

Results from two different experiments based on entropy model have also been included. Enron dataset has been used for the evaluation of first version of the model. In the graph presented, nodes have been used to represent actors or individuals and edges to represent actions, actions are defined as any Phone call, email or meeting. Actors have been divided into three categories namely leader, followers and the mediators by the authors. Comparison with betweenness centrality approach has been presented and it has been claimed that betweenness centrality and approaches like that lead to poor answers when our network consist leaders and followers.

2.6.2 Game Theory

All methods that we have seen in the previous sections basically belong to Network Sciences. A lot of work has been done in the area of social network analysis to detect the most influential actors in a network. More recently game theory has also been introduced to support in the progress of centrality measures in network analysis [32]. Games theory deals with models of competition and cooperation and studies situations where players can get benefits from working together. This concept has been compared with formation of covert networks in [33]. Members of a terrorist network also work as players together to achieve an objective. Members of covert networks have to perform different tasks in order to perform a terrorist event. So different members perform different tasks which are synchronized through communication in order to achieve the target. Because of this reason, covert network rely on communication network to do such acts of recruitment and planning [34]. So to only focus on how these members can operate in optimal way is not important but also how power is allocated among them is important. Game theory allocation rules can be used to analyze power allocation of members of a network [35].

According to [33], cooperative game theory can enable us to model terrorists engaged in coalitions. So by combining cooperative game theory concepts with graph theory, it becomes possible to define centrality measures that are more close to mirror hidden groups structured according to network topologies. For example in drug trafficking networks, values for coalitions can represent the amount of money that respective coalitions obtain by cooperating [33]. An analysis of such networks by applying connectivity games can help detect key players. Authors have used a cooperative game theory solution concept named shapley value to get aid in refining centrality measures for covert networks. Another good related work has been done in [36] where the same concepts of shapley value has been applied to find out the top-k nodes in a network.

2.6.3 Use of Behavioral Profiles

Use of semantic graphs to create behavioral profiles of individuals of a group and then using them to detect key players is also a new idea applied in this field.

A semantic graph is used to represent semantic structures in terms of nodes and relationships between them. A semantic structure is a structure in which two nodes are connected through at least one relation. A node can be any type of object like an individual performing an activity or a

location or even an event itself and the link is the relation between any two such nodes. In semantic graphs, relationship can be between any nodes that may not necessarily of the same type e.g. relation between an individual and a location, and also there can exist more than one relation between two nodes so because of this property semantic graphs are also known as Multi Relational Networks.

A un-supervised framework for profile generation named SoNMine has been introduces in [37]. Profile generation for every node in done on the basis of relations it in involved in. In the work presented by S. Karthika et al in [37], the profile generation uses the semantic graph as its input. A semantic graph uses graph to represent semantic structure in terms of nodes and relations between them and after semantic profile generation (described as a collection of condensed paths generated through variable relaxation approach), the behaviors of nodes are analyzed. Outlier detection is done then on semantic profiles and the nodes who communicate highest are considered the Key Players.

2.7 SNA in terrorist networks

Social Network Analysis (SNA) as discussed in the previous chapter has been applied to a number of fields where ever a social structure can be found. Because of having a social structure SNA has also been applied to the field of combating terrorism like other similar fields. Terrorist networks appear as organizations with some hidden structure and hidden objectives. Krebs, a very famous SNA researcher has quoted in one of his very famous papers [6] some work done by social network theorists who had studied cover, illegal or secret networks as terrorist networks fall under this category. He has quoted work done by Malcolm Sparrow in 1991 who had an excellent overview of SNA to criminal activity. He mentioned three problems of analyzing such networks;

- the first one is incompleteness i.e. the predictability of missing individuals or interactions that analysts will not uncover.
- The second is the fuzziness of boundaries, i.e. the difficulty of deciding who is in and who is out for the network.
- The third one is the dynamic nature of such types of networks. Such networks are not static, they remain changing.

So sparrow suggests looking at waxing and waning strength of a tie depending upon time and task in hand rather than focusing on presence or absence of a link or tire between two individuals. Two other researchers Wayne Baker and Robert Faulkner (Baker and Faulkner 1993) recommend extracting relationships from archival data.

Bonnie Erickson (Erickson 1981) reveals that prior contacts between the members of such groups are of more importance. According to him members of such groups are linked together through very strong ties which are not easily visible like kinship, training camp fellowships and so on.

Krebs has worked on terrorist network of 9-11 and concluded his work as to sketch a correct depiction of a covert network; task and trust ties between the conspirators need to be identified. He suggests the same four measures that are mapped in business organizations i.e. trust, tasks, money & resources and strategy & goals can be used to get information out of illegal networks. Some of the well known contributions made in the area of applying SNA in the field of counter terrorism are presented in the following sections.

2.7.1 Detection of Chain of Command in Terrorist Cells

Jonathan D. Farley contributed with his new mathematical approach to destabilize terrorist networks [38]. Jonathan was of the idea that modeling terrorist organizations as graph is not enough to have enough knowledge about it especially if objective is to deal with the threat. According to him if a terrorist organization is modeled as a graph, an important aspect of the structure i.e. the hierarchy of that network and the fact that they are composed of leaders and followers is ignored. Jonathan proposed an alternative technique that depicts the organization's hierarchy in a better way. In his work, a diagram of an ordered set is generated that consists of leaders which are represented by the topmost nodes and foot soldiers which are depicted with the bottom nodes. To destroy a network, chain of command is recommended to be destroyed in which order flow from the leader to the foot soldiers who are actually the fighter, or the implementers e.g. actual hijackers or the suicide bombers or the assassinators in terrorist networks.

Mainly researchers have focused on how using structural properties the network can be destabilized i.e. analyzing the network, which nodes we can remove to break up the network into

smaller non communicating networks. This work is different from this traditional approach because Jonathan worked in order to detect the leaders and the followers so that the leaders may cut off from the followers and if such is done, the network can be claimed to be neutralized.

The reason behind this idea is, it is not feasible to capture each and every member of terrorist cell in order to prevent a terrorist event who may be also separated by geographical locations, may be residing in different countries even. It may also not even be feasible and cost effective to capture a majority of the members. So an optimal solution is to find the n number of nodes whose removal can neutralize the network preventing a future terrorist event.

Jonathan claims that his ideas should enable intelligence agencies to state that they are for example 85% sure that they have broken the terrorist cell that was target. The definition of broken is debate able. But it is also recognized that if say for an example intelligence agency is 85% sure, there still remains a chance of 15% that the network has not been broken and can still function. So summarizing Jonathans work, the basic idea is to use ordered pairs to model complete terrorist cell in terms of leaders and followers ranging from the leaders to the foot soldiers. Such modeling will clearly give intelligence agencies information about the chain of command and once chain of command is known, cell can be prevented by destroying the chain of command by killing or capturing the leaders or other important nodes.

2.7.2 Matrix Decomposition

As an aid to social network analysis, the use of matrix decompositions to extract more information from a graph to model a terrorist network has also been found useful [39]. Three types of matrix decompositions can be used:

- Singular value decomposition is normally used for dimension reduction, in [39] same has been used for a tool for graph partitioning and also as a way to find out more anomalous hence more important nodes.
- Semi discrete decomposition is used to cluster data. So it is a clustering tool in the whole process
- Independent Component Analysis partitions data graph to components that are the most like cliques (sub groups) as possible.

Matrix decomposition in terrorist networks analysis avoid the weaknesses of traditional link analysis by using extra information of both edges and nodes. In the work of D. B. Skillicorn published in [39], matrix decomposition method has been applied to analyze the Al Qaeda network.

2.7.3 Dynamic Network Analysis:

Most people have a perceptive understanding of hierarchies of normal organization but covert organizations like terrorist organizations have different network structures. Being cellular and distributed is a key property of such networks. So these networks cannot be treated as ordinary organizations' networks because of having so dynamic in nature. Bases on these facts the concept of Dynamic Network Analysis (DNA) has been introduced in [40]. DNA is an advancement which has been introduced as an extension to traditional methods in order to incorporate handling of dynamic nature of terrorist networks. DNA has three key advancements:

- The Meta-Matrix
- Treatment of links as variables so giving capability of having associated weight or probability
- Combining social networks with cognitive science and multi agent systems to enable agents to adapt.

Because of meta matrix a set of networks connecting various entities like individuals, knowledge, resources etc are combined to describe and predict system behavior. With variable link feature links are analyzed as ranging in links' likelihood, strength and direction rather than being simple binary connections just representing presence or absence of a link. The use of multi agent network models enable user to reason about the dynamics of complex adaptive systems [40].

Over all idea is to calculate the performance of the system, then by using any importance criteria like any centrality measure or else find the key player, eliminate them and then check again the performance of the network in order to measure the degree of destabilization.

As a proof of concept, DNA has been tested using data collected on as embassy bombing in Tanzania.

2.7.4 Investigative Data Mining:

Investigative Data Mining (IDM) provides the capability to map a covert cell and also to measure the specific structural and interactional criteria of such a cell [28]. The focus of IDM is to uncover individual's interaction patterns and then to interpret these patterns to predict behavior and decision making within the network. Also the level of covertness and efficiency of a covert cell can be measured using IDM. The level of activity, ability to access others and the level of control over the cell can also be measured. Because of these features, counter terrorism applications can be drawn which can help assessing and neutralizing a terrorist network. Number of variety of investigative data mining can be found in the literature. Link analysis is one of them that use search and probabilistic approaches to find structural characteristic in the network such as hubs, gatekeepers, pulse takers and also identification of potential relationships for relational data mining [41]. IDM link analysis techniques have been implemented and proposed to be used to detect the hidden relationships between members of terrorist organizations. Four case studies have been taken and networks involved have been analyzed in [28]. The case studies taken include Bali Night Club Bombing Terrorist Attack, Dirty Bomb Plot, WTC 1993 Bombing Plot and September 11, 2001 Terrorist Plot.

2.8 Proposed Relative Degree

Relative Degree, A new measure of social network analysis has been purely designed to detect the group leaders in a terrorist network having some special characteristics as observed from real terrorist networks datasets. First role of group leader in a terrorist network and then relative degree in presented in the upcoming sections.

2.8.1 The Role of Leader in a Terrorist Group:

Leader ship is a place where exceptional individuals have been most apparent in the terrorist realm. Scott Stewart wrote in his article published in "*Stratfor Global Intellignce*"

"Although on the surface it might seem like a simple task to find a leader for a militant group, in practice, effective militant leaders are hard to come by." [42]

The reason behind the toughness of finding militant leader in the quality of skills the leaders have which they use both to run the group and also to keep themselves hidden physically and virtually. Leaders of such organizations normally make plans and give instructions to their next

whose further responsibility is to coordinate terrorist activities, pass by instructions given by leader, hire suicide bombers and other personnel and also coordinate trainings and finances etc.

The route of al Qaida's grant in Saudi Arabia is a prominent evidence of the significance of leadership to a terrorist group. Al Qaida was extremely active in Saudi Arabia in 2003-04 under the leadership of Abdel Aziz al-Muqrin [42]. At that time a number of heavy attacks were made inside Saudi Arabia creating a state of panic. The situation was getting worse and according to Scott, it seemed that Saudi Arabia will become next Iraq. However, after death of al-Muqrin in June 2004, the organization began floundering. The next leadership after him lacked confidence and each one proved to be ineffective as Saudi security forces killed many of them.

Because of this failure of leadership, the whole organization became unstable, unorganized and frustrated. Many travelled to places like Iraq and Pakistan to do perform such activities.

Several terrorist who remained in Arabian Peninsula joined Al Qaida's branch in Yemen and formed a new network under the leadership of Al-Wahayshi in Jan 2009, who was the leader of Al Qaida in Yemen and who had served directly under Osama Bin Laden in Afghanistan before being arrested in Iran.

Al-Wahayshi proved himself strong leader because of the increase in operational potential and increased tempo. The activeness of Al Qaida has increased and decreased based on the abilities of the leader. All these facts are clear indication of the importance of leadership in terrorist organizations.

The same article quoted above has been concluded by the following statement:

"Stratfor believes the groups' failures also stem in large part from their lack of effective, dynamic leadership" [42]**.**

Another article says:

"Leadership decapitation has largely failed to produce desired policy results against organizations other than terrorist groups, such as state regimes and drug cartels" [43]**.**

2.8.2 Centralized and Decentralized Networks

A centralized network consists of one or a very small number of central nodes. Such central nodes connect several other nodes of the network indirectly. These central nodes are also known as hubs because of their functioning. To destabilize such networks, removal of such central nodes or hubs have been found very effective as the network can be broken into small dislocated

disjoint cliques which may not be independent to achieve some objective if the central nodes or hubs are eliminated from the network. Hubs or central nodes are the nodes that have more central position as far as the structure of the network is concerned, meaning that the nodes that have higher values of the traditional social network analysis measures like degree and betweenness centrality.

Converse to such central networks, a network which is less centralized has smaller number of break down points. Several points of failure caused by attacks can be tolerated by such networks and they can still survive and function. A typical decentralized network is shown in Figure 2.5.

The structure of a decentralized terrorist network can be derived on the basis of information published in [44].

A leader leads such network who acts as a mentor and provides guidance on how to organize and motivate group operatives.

All the centrality measures that have been proposed in past and are mentioned in Chapter 2 of this thesis detected the most central or hub nodes in order to destabilize the network but we believe that in the light of available literature and ground facts, the role of leader in a terrorist group can also not be neglected. We believe that elimination of a terrorist group leader means the destabilization of that group hence preventing that group to carry out such activities.

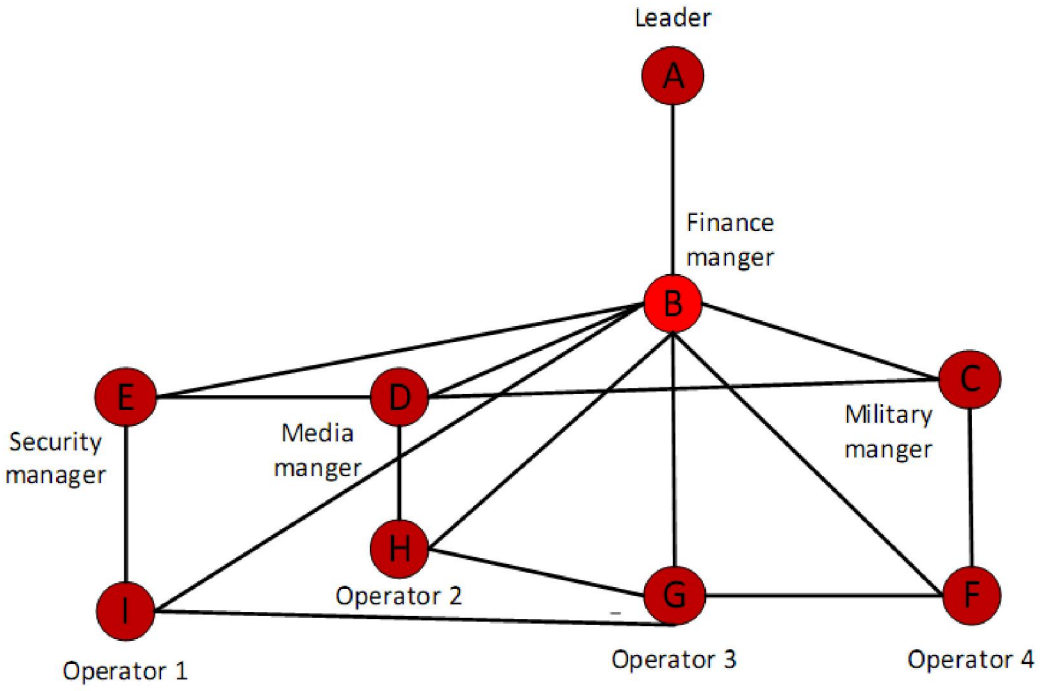

Figure 2.3 A decentralized terrorist network

Figure 2.4 Western intelligence officials believe that the organization, al Qaeda (the base), has a hierarchical structure. Bin Laden, who for security reasons moves constantly around Afghanistan, mostly in the Kandahar region, heads the organization [45]

Leadership Hierarchy of Al Qaeda is shown in Figure 2.4

This significance of leader is the first motivation of the proposed model. In the next sections, the model will be discussed in detail covering ho leader detection portion of the model works.

2.8.3 Relative Degree

Every SNA measure discussed above focuses on a specific nature of nodes and finds the most important node according to that specific feature. For example degree centrality clearly focuses on number of relations. So if degree centrality is considered as a measure to detect the most important node in a network, clearly the person with highest number of directly connected individuals will be taken. If a social networking website is taken as the target network, the number with highest number of friends, if under analysis is a computer network, the most central computer having highest number of network interface and if the network under consideration is a sexy relation network, the individual with highest number of relations will be declared the most important node. Similarly if betweenness or closeness centrality is used, they will work for a specific property.

All existing methods that have been presented in literature have been devised for different types of social networks but all of them consider the node's own position in the network to determine its relative importance. Some also consider neighborhood as Eigen vector centrality as proposed in [29] but all of these ignore its own position and just consider neighbors, so a gap is there for a measure that takes both node's self position and neighbors importance into account. This is the first reason of devising relative degree.

Secondly till now there is no social network analysis measure that has been designed purely for terrorist network analysis keeping in core the key characteristics of terrorist networks which are different from typical social networks. Terrorist network have many differences from a typical social network so traditional SNA measures may not be effective in order to find the most important nodes in a terrorist network reason being covert networks often don't behave like normal social networks [8] because of many reasons. A very basic, commonly obvious and important reason is secrecy. The most important nodes in a terrorist network try their best to hide themselves minimizing the probability of being trapped under magnifier hence preventing the group achieving objective. Keeping this different specification of terrorist network, the new feature "Relative Degree" has been designed and proposed.

A leader in a terrorist group has a very important role in the network that can not be denied as discussed earlier in this chapter but on the other hand because of secrecy, a leader is least connected which is contrary to all traditional SNA measures. Almost all of them look for most connected node which is clearly not the case in a terrorist network. As shown in Figure 2.3, a leader has lowest connectivity but is directly connected with the most influential node in order to run the group having a very strong tie due to continuous involvement in the activates of the group. Hence concluded there are three considerations to detect a leader in a terrorist group in general and a decentralized terrorist group in particular.

- The first is leader node will always have a very less number of connections due to secrecy, as is clear from many examples including the very famous Usama Bin Laden, leading the most famous Al Qaeda hence implying that the value to degree centrality of leader node will be low
- The second is that leader will be connected to the highest degree node at one hop in order to monitor, coordinate, manage and run the group
- The third is that the tie between the leader and the highest degree node will be very much strong.

All these three considerations have been taken to create the relative degree. Relative Degree is defined as "Ratio of degree of maximum degree first hop neighbor to the degree of the node under consideration multiplied with weight of the tie connecting the node under consideration and the maximum degree first hop neighbor". The measure has been designed specifically for weighted networks where link weights represent ties between the actors but can also be applied to other networks as well. Mathematically Relative Degree is given in **Equation 2:12**.

$$\sigma(v) = \frac{Max(Degree(i))}{Degree(v)} * Weight\ (Max(Degree(i)) \qquad \textbf{Equation 2:12}$$

Where i is the set of all first hop neighbors of Node v.

The algorithm for Relative Degree is as under:

```
RELATIVEDEGREE (Node N)
  begin
                max:=0,weight:=1
   foreach (Node m in AdjacentNodes of N)
```

```
    if (m != N)
      if (max < Degree(m))
        max := Degree(m)
                                        weight := weight(m,N)
      End if
    End if
  End foreach
  if (Degree(N) != 0)
    return (max / Degree(N)) * weight
  else
    return 0
  end if
end
```

Consider the network shown in Figure 2.5 A simple network consisting of 9 nodes and 9 edges. Clear from figure Node 'C' has the highest degree i.e. the maximum number of adjacent edges i.e. 5. If we apply degree centrality on the following network, Node 'C' will be detected as the most influential node. This measure is useful for typical social networks but it may not work for a covert network. If we apply the proposed Relative Degree on the same network, starting from node A, A has three neighbors, B, C and G. Among its three neighbors, C has highest degree i.e. 5 and Degree of Node A is 3 so relative degree is given by 5/3 * 1 = 1.67. For simplicity assume that all links have weight of 1.

So applying relative degree on all the nodes, Node F is detected to be the most important because Node F has one adjacent node i.e. Node C with degree 5 while own degree of Node F is 1, therefore relative degree of Node F is given by 5/1 * 1 = 5.

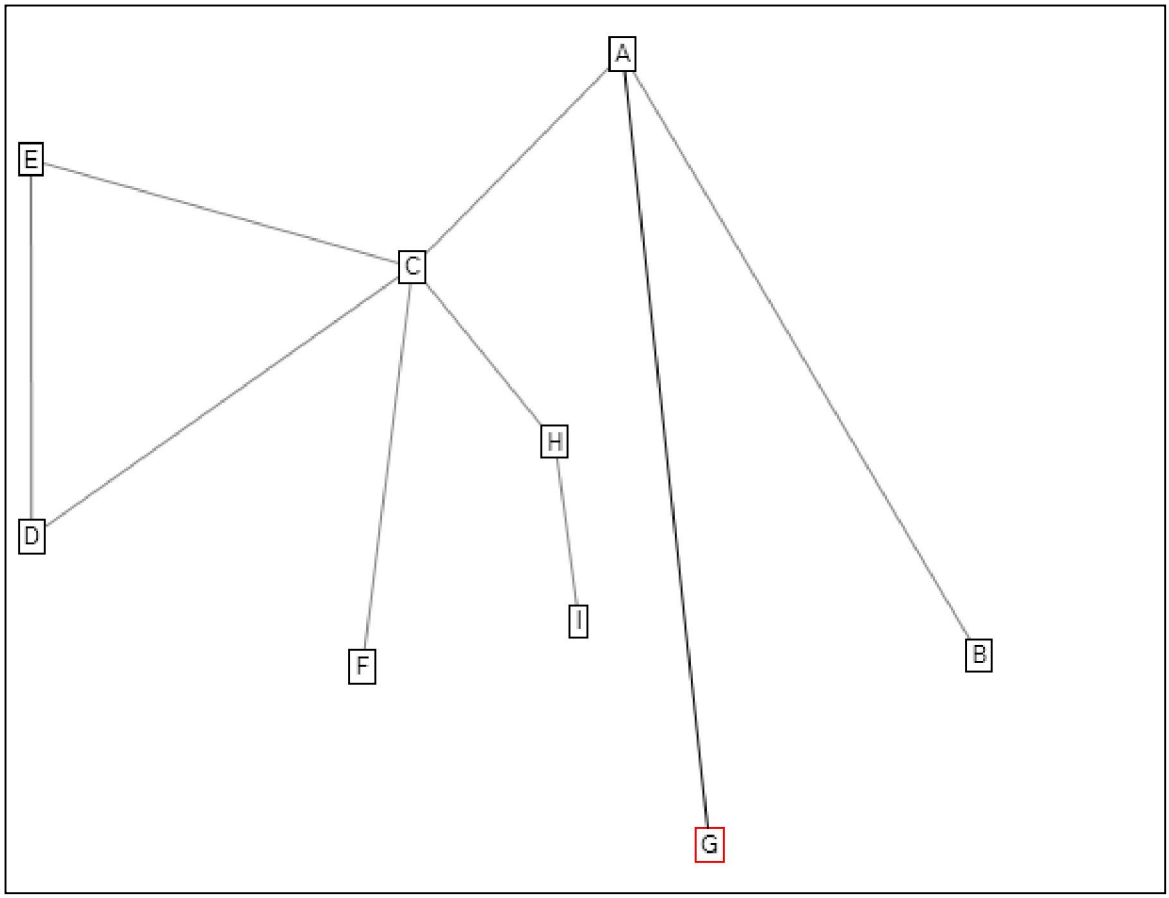

Figure 2.5 A simple network

The reason of detection of Node F as the most important node is aligned with the foundation of relative degree. Firstly; Node F is directly connected to the highest degree node of the network depicting that it can directly send any instructions, orders to the highest degree node which can then disseminate the instructions the desired nodes. Secondly; own degree of node F is low depicting the intention of staying hidden in the network.

2.8.4 How Relative Degree is Different from other centrality measures

As discussed earlier, social network analysis covers the analysis of relations, interactions or links between different actors/individuals/nodes in a social network. One core feature of social network analysis is to find the most influential actors from the network based on the basis of relationships. Previously all the measures included in social network analysis, that have already been discussed above, use mathematical models to analyze the structural position of a node to compute the node's relative importance in the network. For example, betweenness centrality

looks at the broker behavior nodes and is based on the assumption that the most broker behavior node is the most influential node, as far as the structure is concerned. All other measures also have such assumptions. But as discussed earlier, terrorist networks have a different specialty that no other social network has and that is the difference of secrecy.

The lead role playing node in a normal social network may intend its prominence to all other members of the network but same is not the case with a terrorist network. The leader in a terrorist network avoids its exposure that can occur through any medium. Because of this intention in mind, leader remains connected with a very few most reliable nodes. Those most reliable nodes are the most central nodes. If any existing social network analysis measure is applied on such networks, to detect the most central nodes, we get those influential nodes which are directly connected to the leader but they are not leader themselves.

So Relative Degree is proposed as a new social network analysis measure to detect the leaders from such covert networks in which leaders have intentions of hiding themselves. As discussed earlier, all the three considerations that leaders role has, have been taken into account while creating this new measure.

So the main difference between Relative Degree and the existing social network analysis measures is of focus of detection. All traditional measures are aimed to find the most centrally structured node, which are also referred as hubs but on the contrary Relative Degree focuses to find the group leader from a terrorist especially a decentralized terrorist network.

2.8.5 How Relative Degree is Different from other ways of finding most important nodes in a social network

As discussed in the above section, the main difference of Relative Degree and the existing measures is of focus. The focus of existing measures are detection of most central nodes but Relative Degree focuses to find the group leader with the help of central nodes detected by existing measures. Similarly all other methods of detecting the most important actors from a social network that have been discussed above, focus on finding a node which has its own importance which is obviously visible but exposing importance is not a feature of terrorist organizations. So detection of group leader from a terrorist network is not similar to detection of most important nodes from other social networks.

2.8.6 Experiments for Group Leader Detection

Group Leader Detector was implemented in an open source social network analysis tool named NodeXL. The proposed Relative Degree was implemented and integrated in the already existing code of open source NodeXL. A brief overview of the tool is presented below.

2.8.7 NodeXL overview

NodeXL is an extendible toolkit for network overview, discovery and exploration implemented as an add-in to the Microsoft Excel 2007 spreadsheet software. The NodeXL tool adds "network graph" as a chart type to the nearly ubiquitous Excel spreadsheet. The intention of developing the tool is to make network analysis tasks easier to perform for novices and experts [46].
Some salient features of NodeXL are [47]

Flexible Import and Export : Graphs from GraphML, Pajek, UCINet, and matrix formats can be imported and also exported to these formats from NodeXL.

Zoom and Scale Zooming into areas of interest, and scaling the graph's vertices to reduce cluttering is also an attractive feature.

Easily Adjusted Appearance Color, shape, size, label, and opacity of individual vertices can be easily adjusted by filling in worksheet cells. NodeXL can also do it by itself based on vertex attributes such as degree, betweenness centrality or PageRank.

Dynamic Filtering Instantly vertices and edges using a set of sliders can be hidden.

Graph Metric Calculations All well known graph metrics are implemented and can be calculated conveniently like degree, betweenness centrality, closeness centrality, eigenvector centrality, PageRank, clustering coefficient, graph density and more.

And from our perspective the most valued feature is the open source code which can be extended in order to implement customized functionality. NodeXL is based on Microsoft Platform so the source code is available in Visual C# which is a Microsoft.Net based programming language.

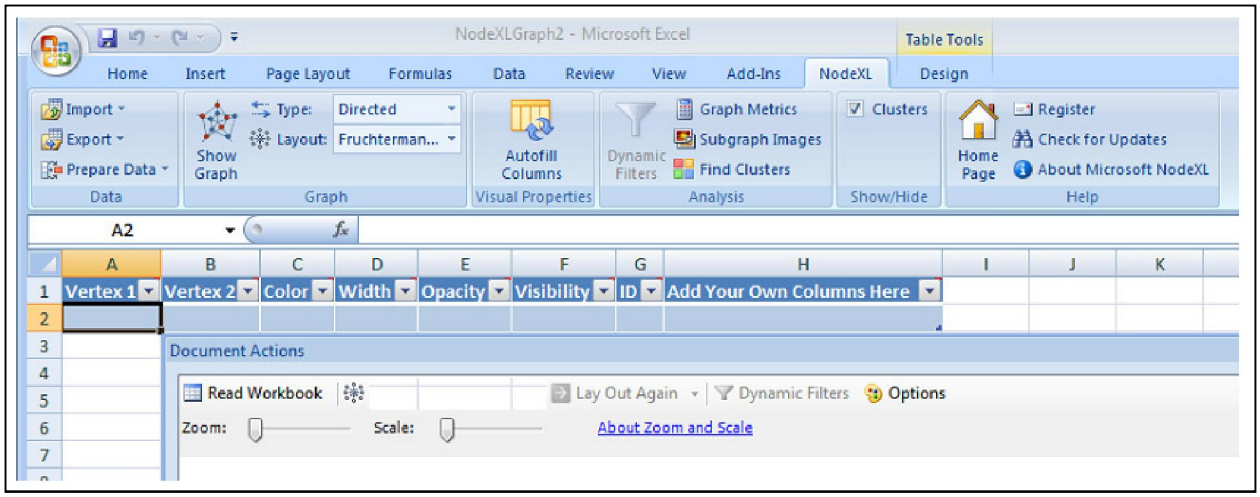

Figure 2.6 NodeXL Menu, Edge List Worksheet, and Graph Display Panel

NodeXL is easy to learn for many existing users of Microsoft Office Excel and has an extendable open source code base. Data entered into the NodeXL template workbook can be transformed into a directed graph chart in a matter of a few clicks. Figure 2.6 is shows a screen shot of the user interface of NodeXL.

The proposed Relative Degree Algorithm was implemented in NodeXL code using MS Visual C#. The extended NodeXL was then plugged with Microsoft Excel and was used for experimentation. Relative degree plugged in with NodeXL is shown in Figure 2.7.

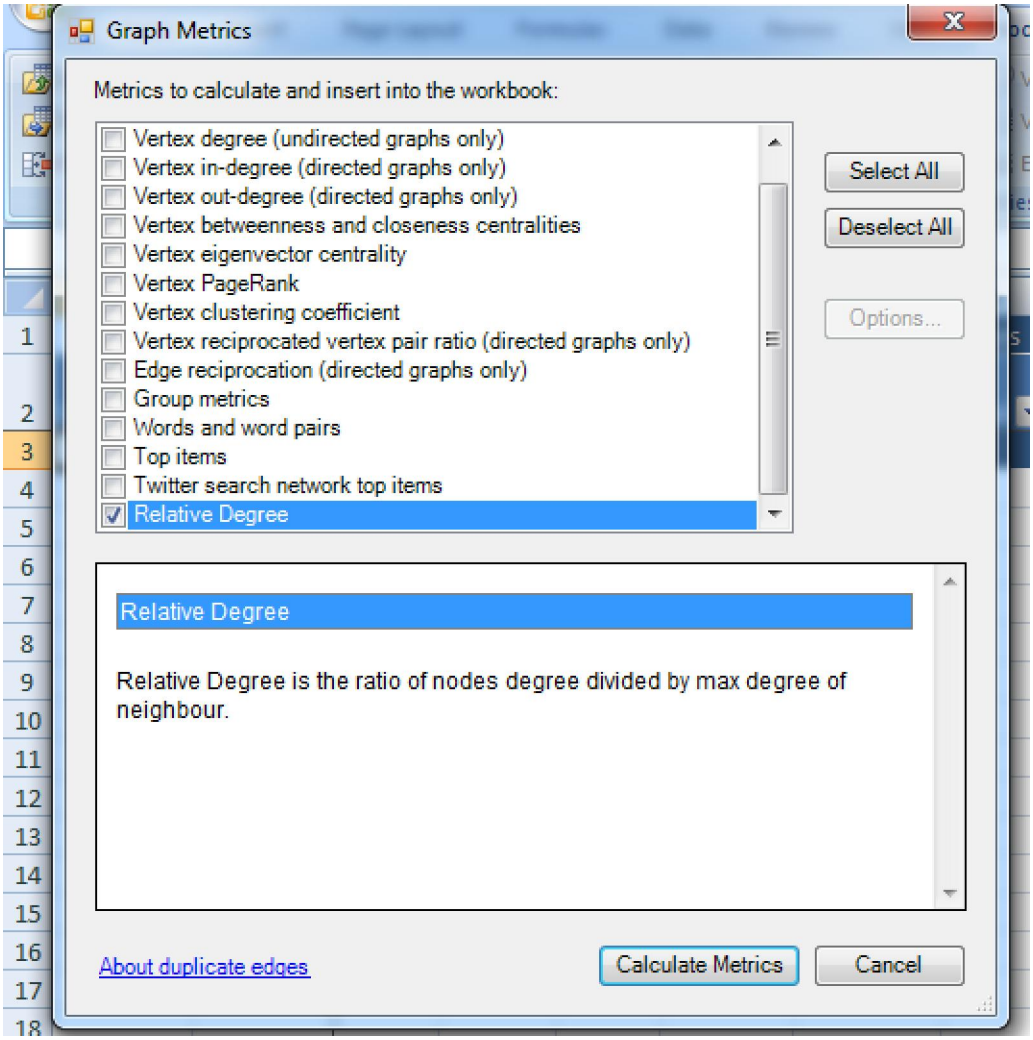

Figure 2.7 Relative Degree implemented in NodeXL Excel Plug in

2.8.8 Case Study 1:

Dataset was taken from a relevant organization containing real data in which roles and importance of every node was known. The dataset consisted of 79 individuals i.e. 79 nodes and 399 edges between them. NodeXL already contains the well known SNA measures implemented in it. Code was extended to incorporate the proposed Relative Degree algorithm. Experiment results are evident of the correctness of the proposed measure. For simplicity weights of all the edges are taken equal to 1. Following table shows the values of well known SNA measures along with value against the newly proposed measure for each individual present in the dataset.

Vertex	In-Degree	Out-Degree	Betweennes s Centrality	Closeness Centrality	Eigenvector Centrality	Relative Degree
A	2	2	0.000	0.004	0.003	4
B	8	8	152.074	0.006	0.022	5.125
C	6	6	57.290	0.005	0.011	2.5
D	7	7	145.833	0.006	0.019	5.8571429
E	1	1	0.000	0.004	0.002	7
F	6	6	54.686	0.006	0.016	6.8333333
G	2	2	0.000	0.004	0.005	7.5
H	15	15	442.581	0.006	0.032	2.7333333
I	2	2	0.000	0.004	0.005	7.5
J	6	6	113.769	0.006	0.018	6.8333333
K	11	11	258.212	0.006	0.024	3.7272727
L	2	2	0.000	0.004	0.003	5.5
M	6	6	33.485	0.005	0.009	1.8333333
N	1	1	0.000	0.004	0.002	7
O	7	7	151.233	0.006	0.021	5.8571429
P	2	2	0.000	0.004	0.005	7.5
Q	4	4	43.613	0.006	0.016	10.25
R	7	7	161.032	0.006	0.015	3.1428571
S	2	2	0.000	0.004	0.002	3.5
T	6	6	40.257	0.005	0.008	2
U	10	10	107.143	0.006	0.028	4.1
V	5	5	23.644	0.006	0.016	8.2
W	3	3	19.941	0.006	0.010	13.666667
X	41	41	2905.094	0.009	0.076	0.5365854
Y	4	4	2.000	0.006	0.016	10.25
Z	22	22	555.013	0.007	0.049	1.8636364
AA	7	7	28.398	0.006	0.021	5.8571429
BB	3	3	0.000	0.006	0.014	13.666667
CC	5	5	0.000	0.006	0.017	8.2
DD	6	6	4.667	0.006	0.019	6.8333333
EE	6	6	5.889	0.006	0.018	6.8333333
FF	2	2	0.000	0.005	0.009	20.5
GG	1	1	0.000	0.003	0.000	3
HH	3	3	286.000	0.003	0.000	0.6666667
II	1	1	0.000	0.003	0.000	3
JJ	2	2	420.000	0.004	0.001	2.5
KK	6	6	75.371	0.006	0.019	6.8333333
LL	2	2	0.000	0.004	0.004	3
MM	2	2	0.000	0.005	0.008	20.5

NN	2	2	0.000	0.005	0.008	20.5
OO	4	4	38.015	0.005	0.007	3.75
PP	2	2	0.000	0.004	0.002	2
QQ	10	10	122.520	0.006	0.028	4.1
RR	4	4	134.120	0.006	0.010	10.25
SS	1	1	0.000	0.005	0.004	19
TT	19	19	566.258	0.007	0.042	2.1578947
UU	3	3	0.000	0.006	0.013	13.666667
VV	5	5	75.921	0.006	0.019	8.2
WW	3	3	1.400	0.005	0.008	6.3333333
XX	5	5	3.536	0.006	0.019	8.2
YY	2	2	1.733	0.004	0.003	3.5
ZZ	5	5	155.315	0.006	0.017	8.2
AAA	1	1	0.000	0.004	0.002	5
BBB	5	5	5.667	0.005	0.013	4.4
CCC	10	10	65.319	0.006	0.025	4.1
DDD	8	8	27.812	0.006	0.025	5.125
EEE	8	8	320.291	0.006	0.017	5.125
FFF	6	6	2.538	0.006	0.018	6.8333333
GGG	1	1	0.000	0.004	0.002	8
HHH	10	10	68.482	0.006	0.024	4.1
III	4	4	0.000	0.006	0.014	10.25
JJJ	5	5	571.223	0.006	0.015	8.2
KKK	7	7	292.677	0.006	0.015	5.8571429
LLL	1	1	0.000	0.004	0.001	7
MMM	2	2	1.733	0.004	0.003	3.5
NNN	12	12	541.331	0.006	0.022	3.4166667
OOO	5	5	5.757	0.006	0.017	8.2
PPP	6	6	37.872	0.006	0.018	6.8333333
QQQ	2	2	0.000	0.005	0.004	6
RRR	1	1	0.000	0.004	0.002	12
TTT	1	1	0.000	0.004	0.002	12
UUU	1	1	0.000	0.004	0.002	12
VVV	4	4	75.253	0.006	0.012	10.25
WWW	1	1	0.000	0.004	0.002	8
XXX	0	0	0.000	0.000	0.000	0
YYY	0	0	0.000	0.000	0.000	0
ZZZ	0	0	0.000	0.000	0.000	0
AAAA	0	0	0.000	0.000	0.000	0
BBBB	0	0	0.000	0.000	0.000	0

Table 2.2 Graph Metric values of members of terrorist group under analysis

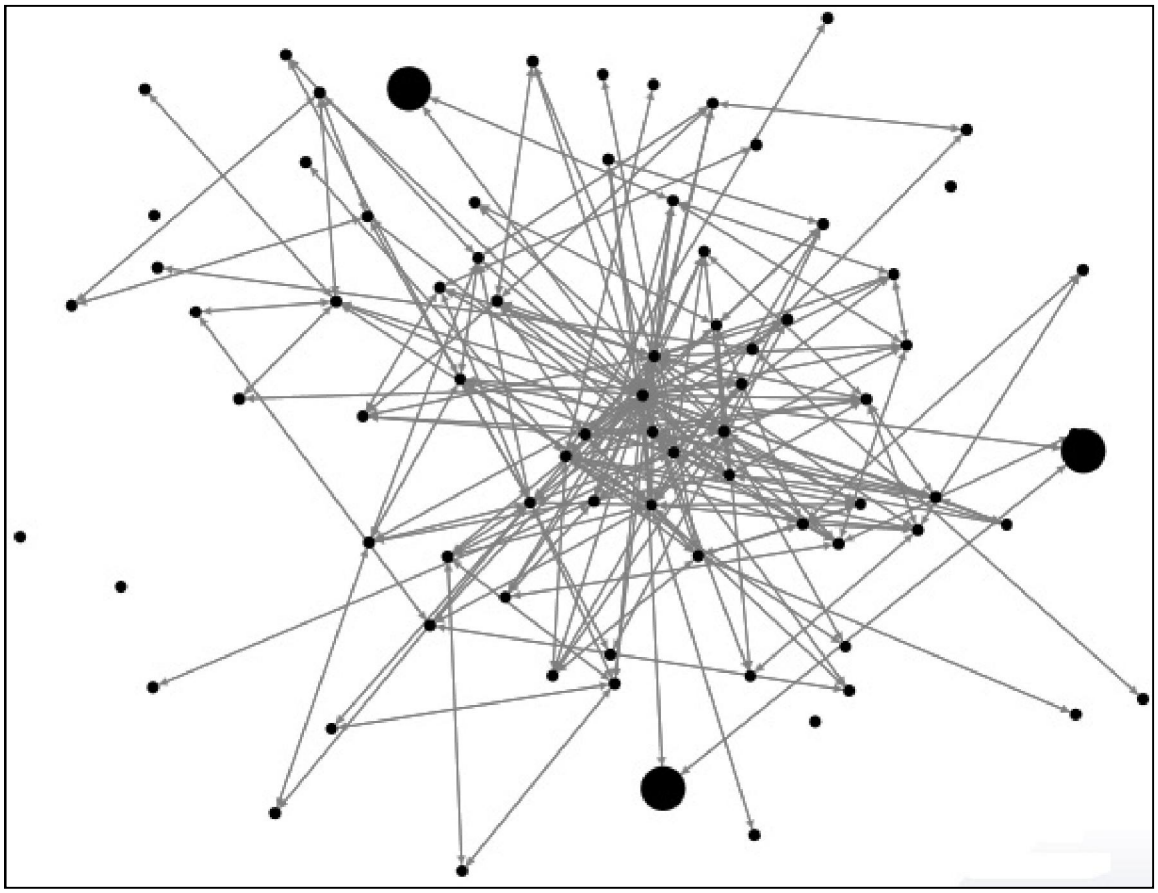

Figure 2.8 Case study 1 Network

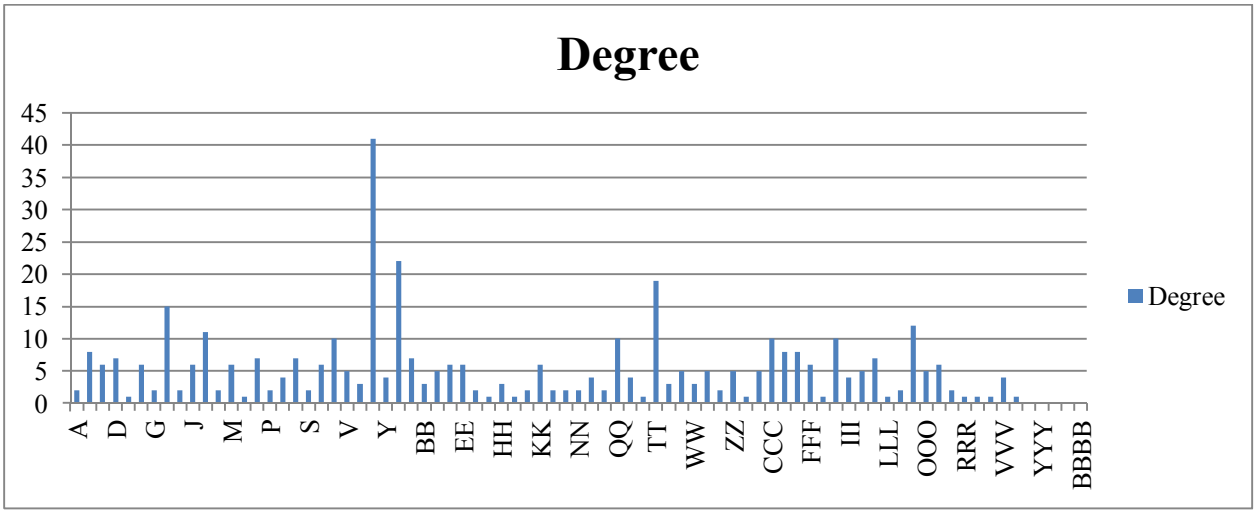

Figure 2.9 Degree centralities case study 1 network

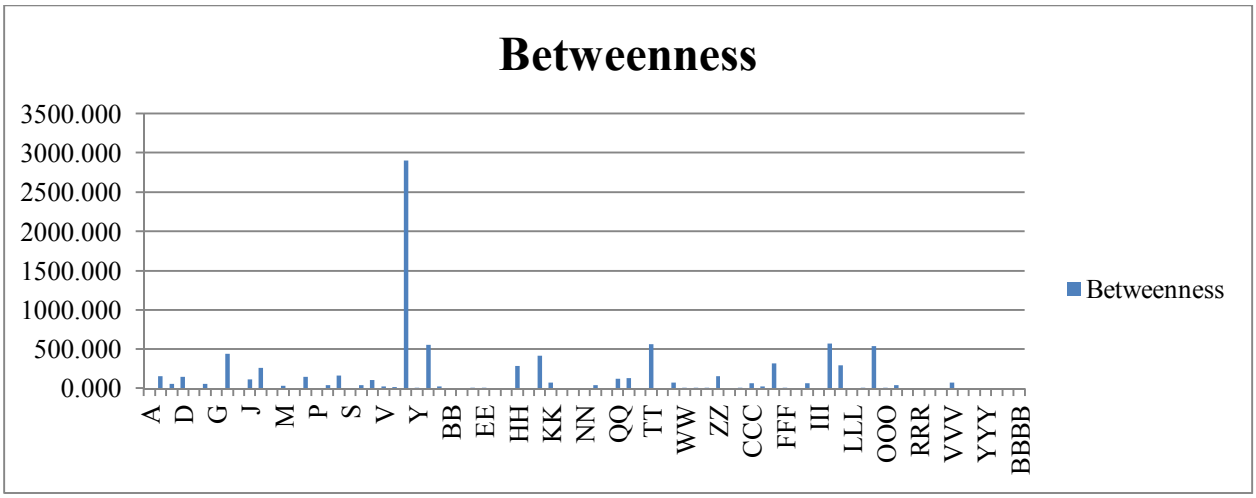

Figure 2.10: Betweenness centralities case study 1 network

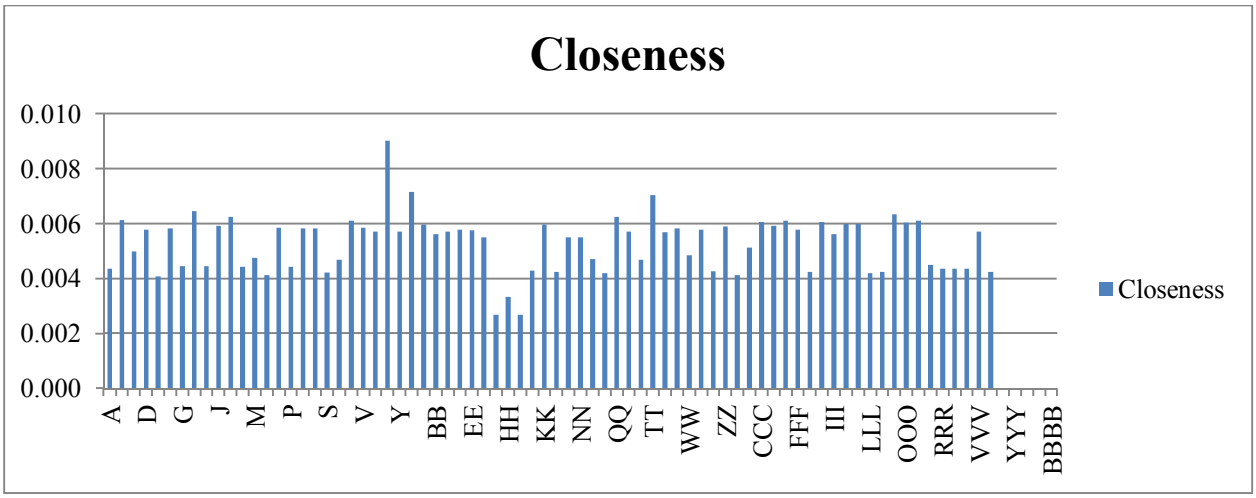

Figure 2.11 Closeness centralities case study 1 network

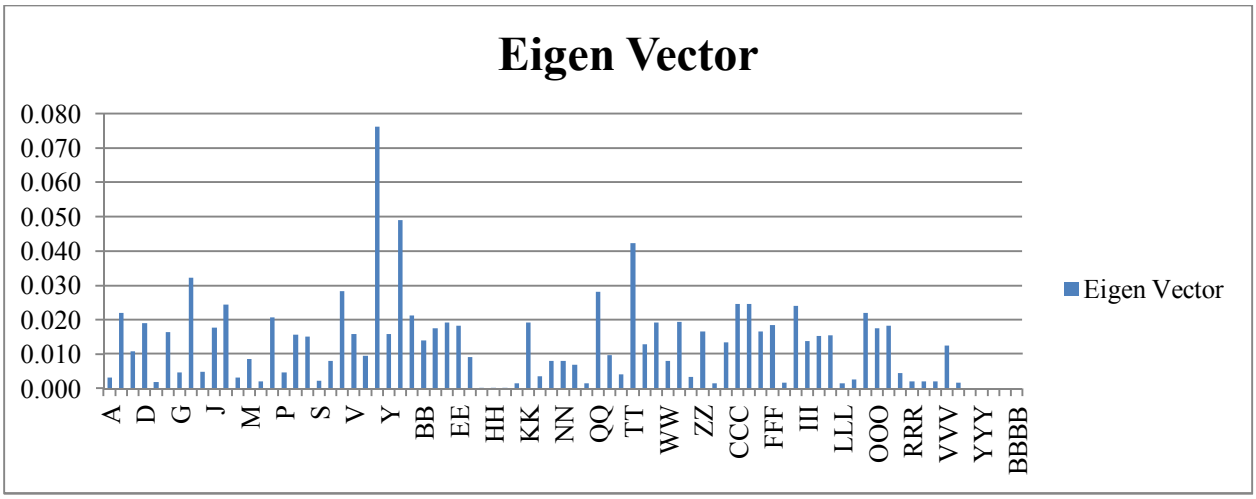

Figure 2.12 Eigen Vector centralities case study 1 network

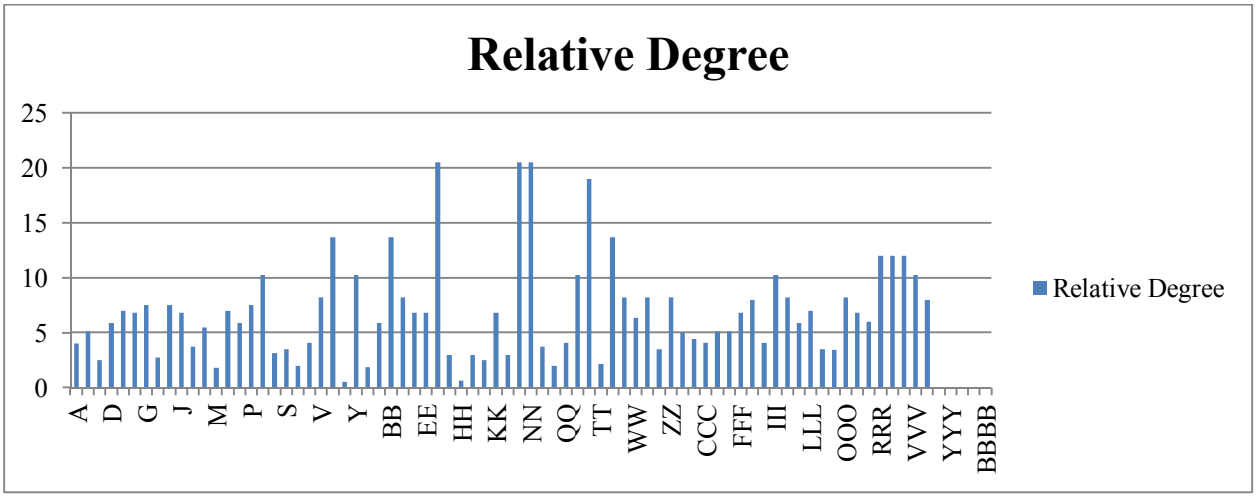

Figure 2.13 Relative degrees case study 1 network

The network diagram of the group under discussion in this case study is shown in Figure 2.8. Figure 2.9, Figure 2.10, Figure 2.11, Figure 2.12 and Figure 2.13 show the plots of values of each member of the group for degree, betweenness, closeness, eigen vector and relative degree respectively.

2.8.9 Discussion on Results

Node 'X' has highest values for all traditional parameters i.e. for Degree, Betweenness, Closeness and Eigen value but has a low value for the proposed parameter. The reason for this contradiction is that all existing parameters have considered only the individual's own influence or position in the network while proposed measure is not totally based on individual's position rather it depends on individual as well as its neighbors importance in the network. So three nodes i.e. node FF, MM and NN have highest value i.e. 20.5 which is because of the fact that all these three nodes have low values for own degree i.e. 2 but they are directly connected to the highest degree node of the network. Having a direct connection to the highest degree node depicts the ease of approaching all other nodes of the network while having own low degree represents the intent of remaining hidden in the network.

2.8.10 Case Study 2:

This case study has been taken from [48]. Description of the data is follows:

On 27th Feb, 2011, five terrorists were killed in Sulaymaniya, Iraq, who were possibly planning an attack on the American university campus in Sulaymaniya. The data set has been created by author of [48] using open source information from newspapers on the internet. The network has been created and modeled in NodeXL. Figure 2.14 shows the network structure of the terrorist network.

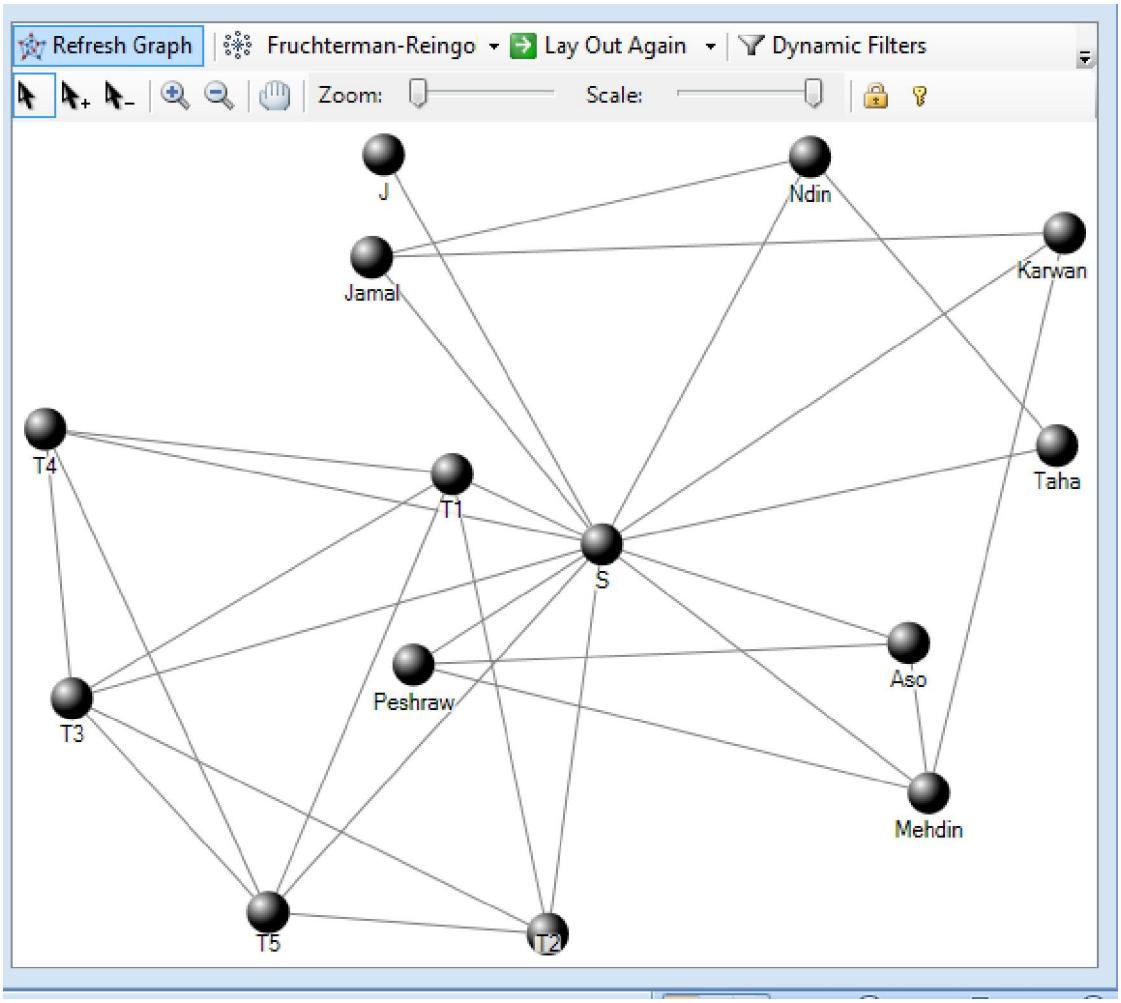

Figure 2.14: Terrorist Network who had possibly planned American university Sulaymania Attack in Iraq

Node	Degree	Betwenness	Closeness	Eigen Vector	Relative Degree
J	1	0.000	0.040	0.030	13
S	13	117.500	0.077	0.158	0.384615385
T1	5	0.500	0.048	0.100	2.6
T2	4	0.000	0.045	0.086	3.25

T3	5	0.500	0.048	0.100	2.6
T4	4	0.000	0.045	0.086	3.25
T5	5	0.500	0.048	0.100	2.6
Aso	3	0.000	0.043	0.050	4.333333333
Peshraw	3	0.000	0.043	0.050	4.333333333
Mehdin	4	2.000	0.045	0.058	3.25
Karwan	3	1.000	0.043	0.050	4.333333333
Jamal	3	1.000	0.043	0.048	4.333333333
Ndin	3	1.000	0.043	0.046	4.333333333
Taha	2	0.000	0.042	0.038	6.5

Table 2.3 Graph metrics of terrorist network

Table 2.3 shows the values of traditional as well as the proposed measure in order to validate the detection of group leader.

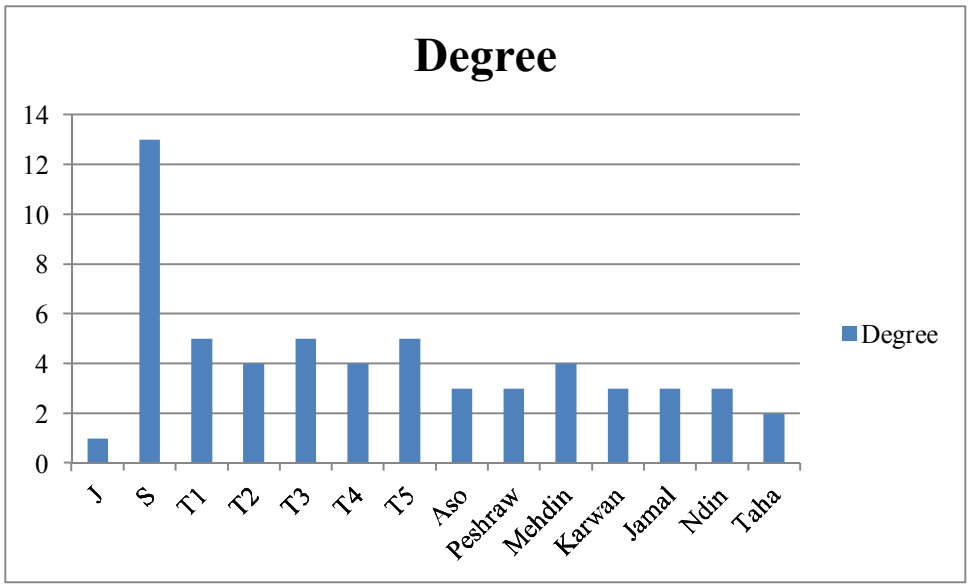

Figure 2.15 Degree centralities of case study network

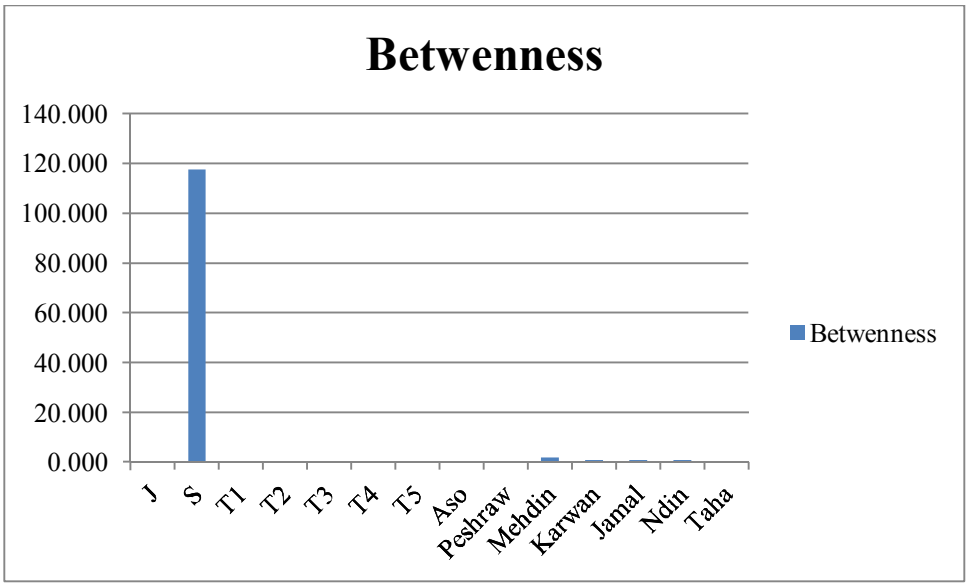

Figure 2.16 Betweenness centralities of case study network

Figure 2.17 Closeness centralities of case study network

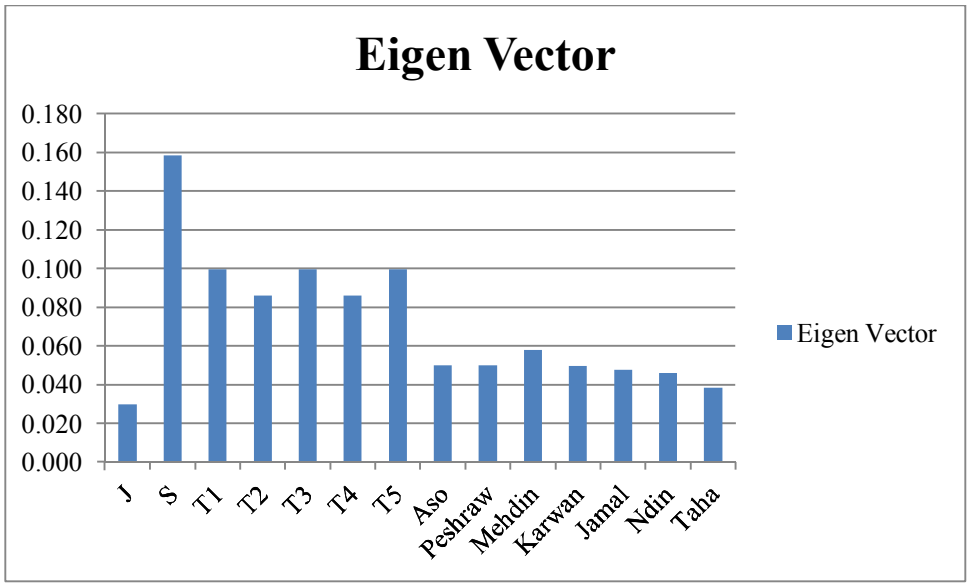

Figure 2.18 Eigen Vector centralities of case study network

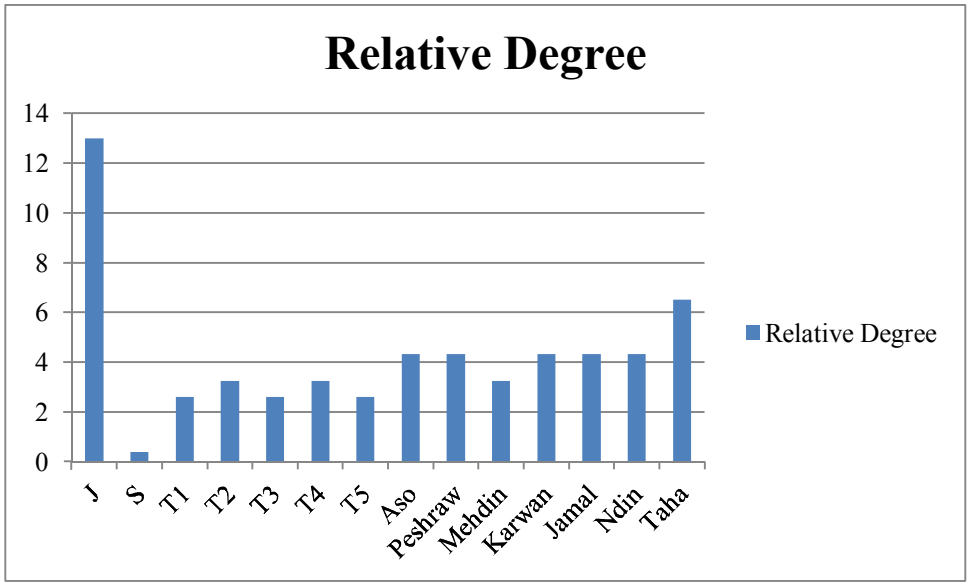

Figure 2.19 Relative Degrees of case study network

For each member of the group under discussion, the plot of values of degree, betweenness, closeness, eigen vector and relative degree and shown in Figure 2.15, Figure 2.16, Figure 2.17, Figure 2.18 and Figure 2.19 respectively.

2.8.11 Discussion

In the results presented in [48] , it has been mentioned that Node S was working as finance manager of J. This implies that J was leading the group. Results shown depict that all traditional measures gave a largest value to node S, who was the finance manager, because of having the most central position in the network. The proposed Relative Degree gave highest values to J, who was actually leading the group. The detection of original group leader validates the correctness of the proposed measure.

2.8.12 Case Study 3:

This case study 3 is about the most famous Al Qaeda leadership hierarchy. The dataset has been created from the information present in [49].

The network created from the information given in [49] is shown in Figure 2.20.

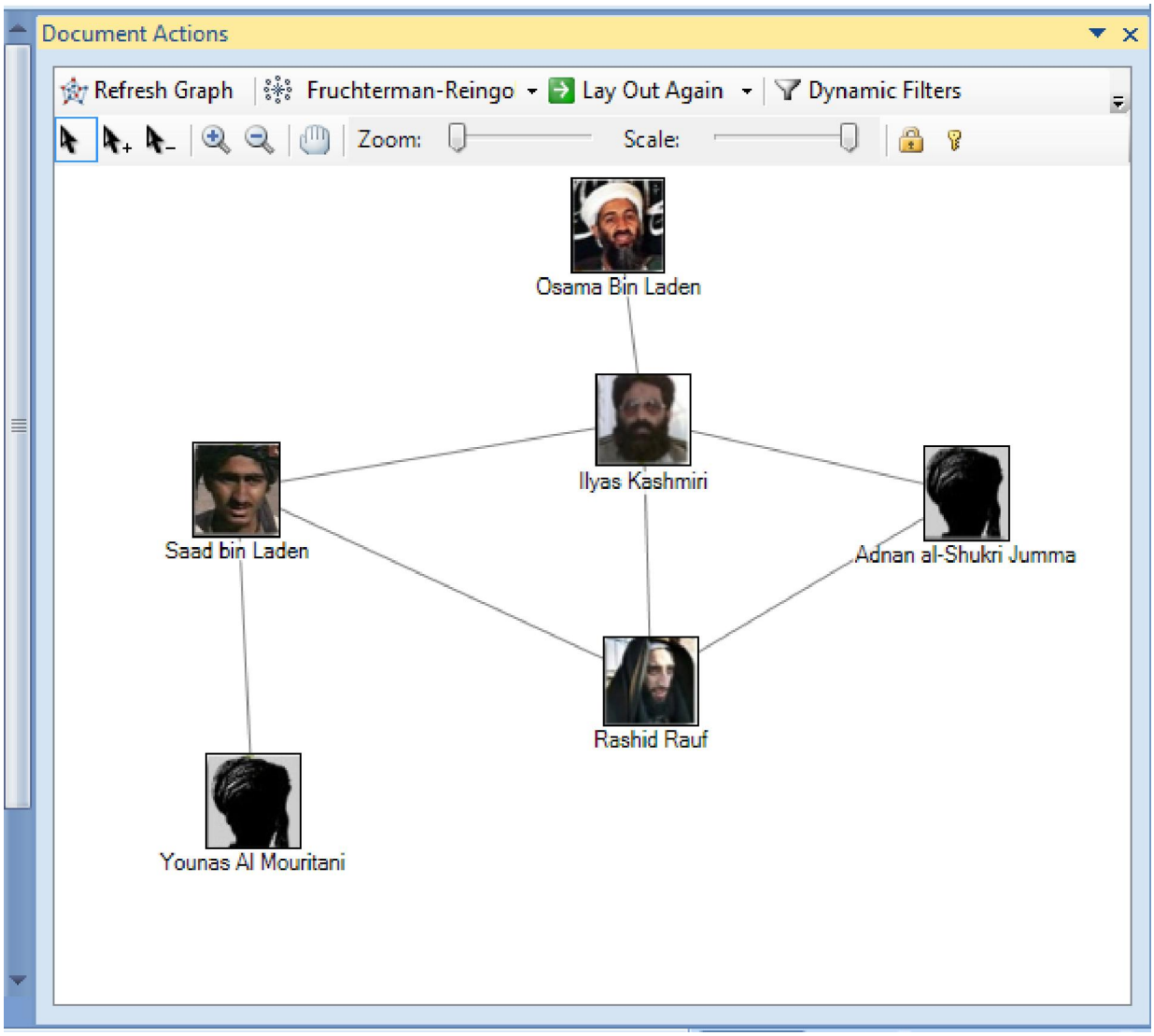

Figure 2.20 Al Qaeda leadership network

Node	Degree	Betwenness	Closeness	Eigen Vector	Relative Degree
Osama Bin Laden	1	0.000	0.100	0.090	4
Ilyas Kashmiri	4	10.000	0.167	0.247	0.75
Adnan al-Shukri Jumma	2	0.000	0.111	0.171	2
Saad bin Laden	3	8.000	0.143	0.197	1.333333333
Rashid Rauf	3	2.000	0.143	0.223	1.333333333
Younas Al Mouritani	1	0.000	0.091	0.072	3

Table 2.4 Al Qaida leadership network

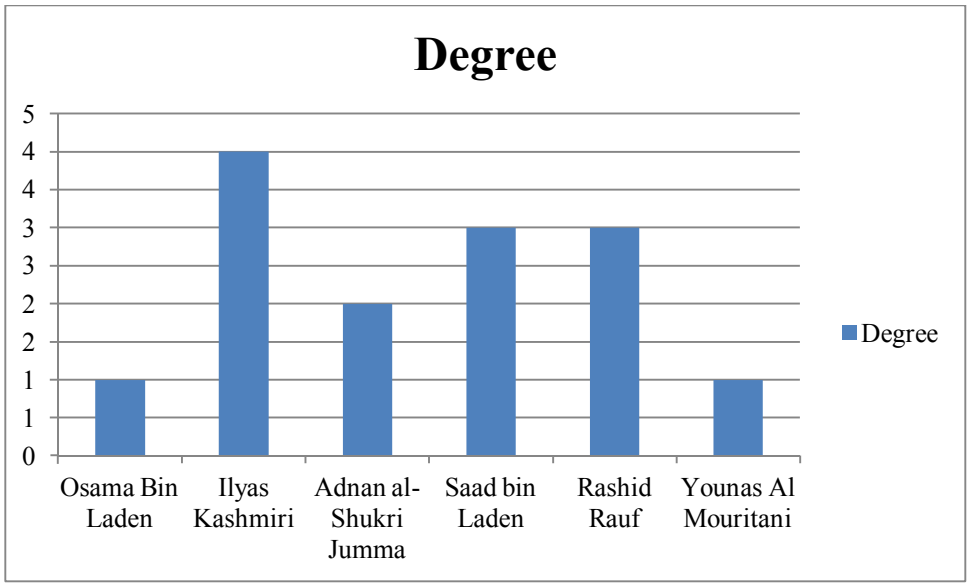

Figure 2.21 Degrees of Al Qaeda Leadership network

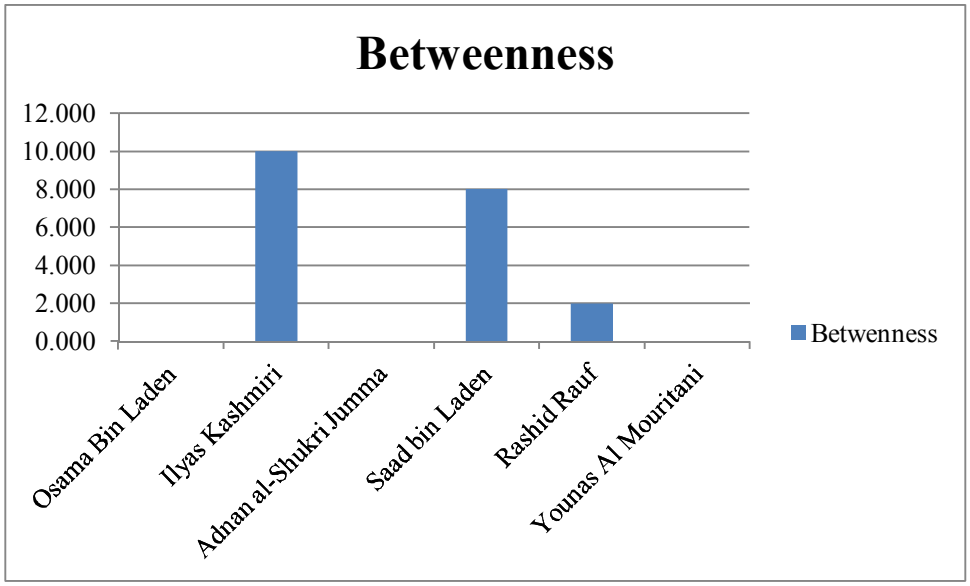

Figure 2.22 Betweenness centralities of Al Qaeda Leadership network

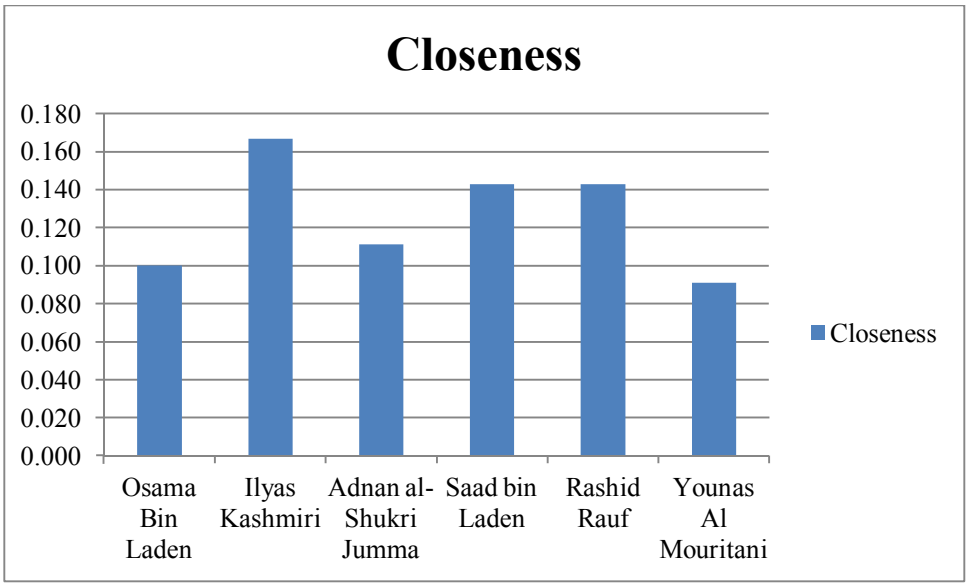

Figure 2.23 Closeness centralities of Al Qaeda Leadership network

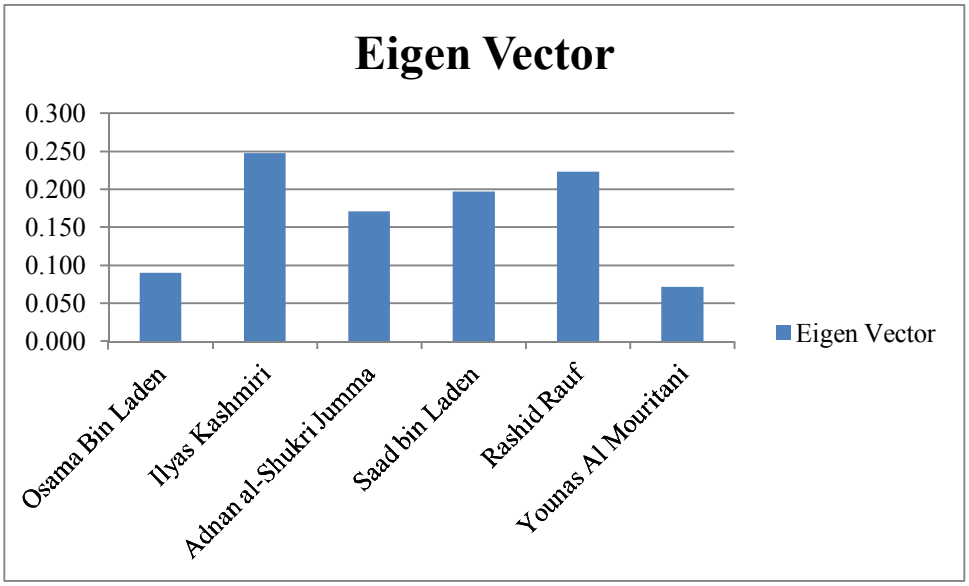

Figure 2.24 Closeness centralities of Al Qaeda Leadership network

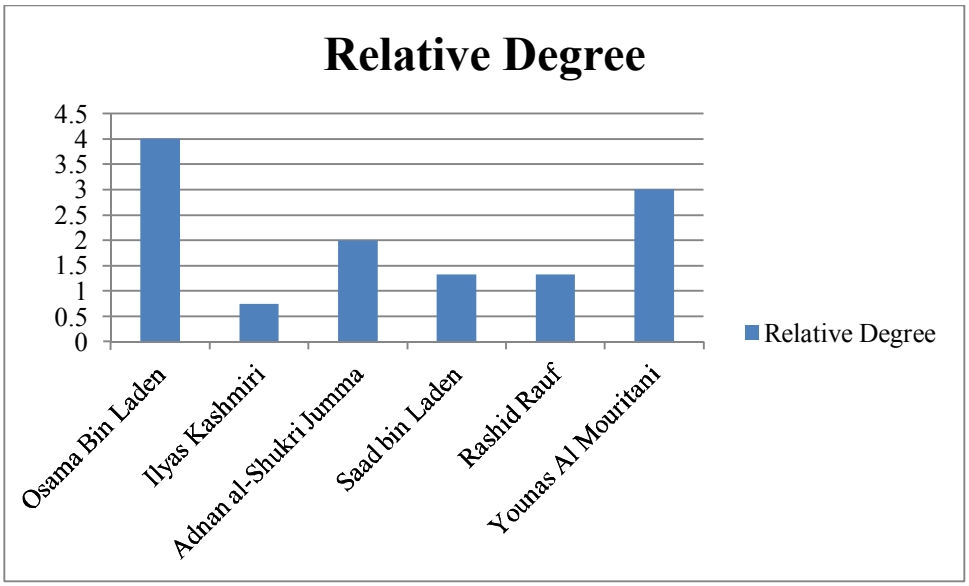

Figure 2.25 Relative Degrees of Al Qaeda Leadership network

Figure 2.21, Figure 2.22, Figure 2.23, Figure 2.24 and Figure 2.25 represent the plots of values of each individual of the network under study for degree, betweenness, closeness, eigen vector and relative degree respectively.

2.8.13 Discussion

In the case study 3 data set, Ilyas Kashmiri is known as chief of 313 brigade so is the most central node in the hierarchy and has interaction to all the lower tier leaders. Because of this reason the values of degree, betweenness, closeness and eigen vector centralities of Ilyas Kashmiri are highest. On the other hand, Osama Bin Laden was the overall leader but he didn't kept contact with any other than Ilyas Kashmiri in the network in order to control whole network through Ilyas Kashmiri and keep himself hidden from being traced. The value of Relative Degree from Osama Bin Laden is highest; clearly declaring him the group leader. Case study 3 also is a proof of validation of the proposed measure.

Chapter 3 Outlier Detection for Event Prediction

Outlier or anomaly is an exceptional data instance value or reading. An outlier deviates so much from normal instances that it depicts something abnormal that can be an error or something like that. Outlier detection has been proposed and applied for detection of an un foreseen event in this chapter. A background of outlier detection, methods and uses existing in literature and proposed use along with implementation and results is presented in this chapter.

3.1 Outlier Detection

The use of outlier detection is to detect data objects or groups of data objects from a data set that are exceptional when compared with the rest of a large amount of data. A definition of outliers is [50]:

"An outlier is an observation that deviates so much from other observations as to arouse suspicions that it was generated by a different mechanism"

Outlier detection deals with detection of patterns from data which do not match to expected normal behavior. These anomalous patterns are often known as outliers, anomalies, discordant observations etc in different application domains. Outlier detection is a well researched area having an immense use in a wide range of applications like fraud detection, insurance, intrusion detection in cyber security, fault detection in security critical systems, military surveillance for enemy activities and so on.

Outlier detection is very important because of the fact that outliers in data point towards something important that cannot be ignored. For example if extra ordinary traffic pattern is observed in a computer network, which obviously is an outlier, could point that a hacked computer is there in the network which may be sending important secret data outside the network. So indication towards a problem is due to outlier detection. Similarly outlier transactions in credit card data could indicate that the card has been misused.

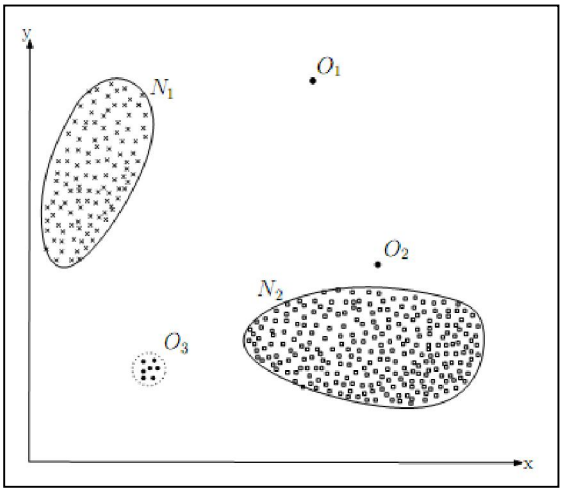

Figure 3.1 Outliers in a two dimensional dataset [51]

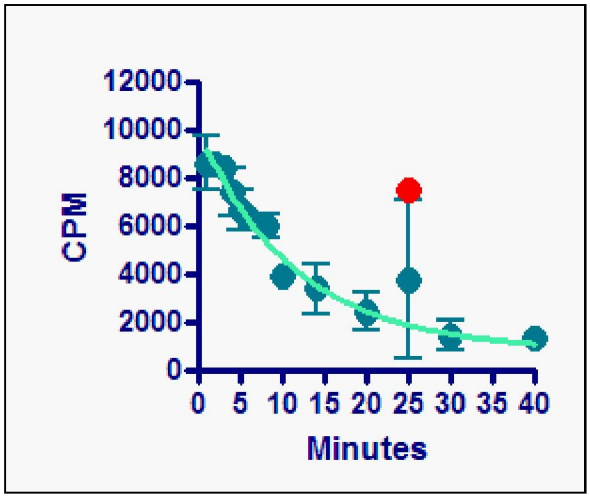

Figure 3.2 A simple example of outliers in a 1-dimensional dataset [52]

Outliers in a simple two dimensional dataset are shown in Figure 3.1. N1 and N2 are two normal regions while O1 and O2 are two outlying instances and O3 is an outlying region. Outliers can be found in almost every dataset that is real. Similarly outliers in one dimensional data are shown is Figure 3.2

Some major causes of existence of outliers include [51]:

- Malicious activity: Any activity that is exceptional and anomalous in a system e.g. credit card or telecom fraud, cyber intrusion, a terrorist activity

- Instrument error: Some outliers may occur due to fault in components of measuring machines
- Environment Change: Such as climate change, new buying pattern among consumers, mutation in genes
- Human error: Data reporting error

Different aspects of outlier detection and presented in the following next section.

3.2 Nature of Input Data

Nature of input data has a key role in any outlier detection technique. Input is normally dealt as a collection of data instances or objects (sometimes also referred as record, point, vector, pattern, event, case, sample, observation, entity) [53]. Data instances are described using a set of attributes (also referred to as variable, characteristic, feature, field, and dimension). The attributes can be of different types such as binary, categorical or numerical consisting of only one attribute or multiple attributes. In outlier detection technique design, nature of attributes has a major role.

3.3 Data Labels

Data labels are used and associated with data instances to declare them normal or outlier. Normally labeling is expensive and also it is prohibitive to obtain labeled data which is accurate and also representative of whole behavior. Normally labeling is done manually by human experts so considerable effort is required to produce a labeled training dataset. Therefore it is more complicated to obtain a labeled set of outliers that covers whole complete behavior.

One of the following modes can be chosen as mode of operation of outlier detection techniques on the basis of the extent to which labeled data is used [54]

3.4 Supervised outlier detection

Supervised mode techniques assume the availability of training dataset having all labeled data i.e. normal as well as the outlier instances. The normal approach is to create a predictive model for normal and the outlier classes. An unseen data object is compared against both the classes in order to decide whether it is a normal or outlier data object.

Techniques operating under supervised mode are highly accurate provided the labels used in the training dataset are accurate. The disadvantage of such techniques is the pre requirement of labeled training dataset.

3.5 Un Supervised outlier detection

Labeled data for training is not required by the outlier detection techniques which fall under this category so such techniques are more widely used. The base behind working of these techniques is the assumption that normal instances are far more frequent that outliers in the given data set. The major disadvantage of these techniques is a high rate of false alarms when number is outliers is higher in the given data set.

3.6 Outlier Detection Techniques

Outlier detection techniques that may be used for assigning grades to different data objects on the basis of their frequency in the given data set are discussed in this section.

3.6.1 Statistical Outlier Detection Techniques

Statistical outlier detection techniques are based on the principal which is: 'An outlier is an observation which is suspected of being partially or wholly irrelevant because it is not generated by the stochastic model assumed' [53], [54].

Outlier detection techniques which fall under this category develop statistical models from the given data and apply a statistical inference test to decide whether the given instance belongs to the model or not. The objects which have low probability of having association with the developed model are labeled as outliers.

Such techniques generally consist of two phases, first one is the training and the second one is the testing phase as presented below:

- **The Training Phase**
  In the first phase i.e. the training phase, statistical model is fit to the given data. Two techniques named parametric and non parametric are used in this step. First type of techniques assume knowledge of the under lying distribution and estimates the

parameters from the given data [53], [54]. While non parametric techniques do no assume any such knowledge [54]. These techniques are normally robust if dataset contains small number of outliers so can work in an un supervised mode. Probability density for both normal and outlier data objects is estimated by supervised techniques while a statistical model which fits the majority of observations is determined by unsupervised techniques.

- **The Testing Phase**

  After the probabilistic model is developed and known test can be applied against any data object to determine whether that is normal or an outlier instance with respect to the developed model. A simple approach is to measure the distance of the given data object from the estimated mean and then to set any point above a certain limit to be an outlier [53], [54].

3.6.2 Clustering-Based Outlier

Cluster analysis [55] is a well known unsupervised machine learning technique which combines similar data objects into clusters. Outlier detection and clustering seem to be primarily dissimilar from each other even though numerous outlier detection techniques based on clustering have been developed. The key assumption in the basis of such techniques is that normal data objects belong to large and intense clusters, while outlier data objects either do not belong to any cluster or form very small clusters.

Most techniques based on clustering find outliers as the by-product of a clustering algorithm [56]. In such algorithms any data object which does not fit into any cluster is declared to be an outlier. The only problem to these techniques is that they aim mainly to find clusters, so they are not optimized to find outliers, and most of the techniques are limited to numerical data.

3.6.3 Nearest Neighbor Based Outlier Detection

Analysis of nearest neighbors is a very well known concept of machine learning and data mining. In nearest neighbor analysis a data object is analyzed with respect to its nearest neighbors [57].

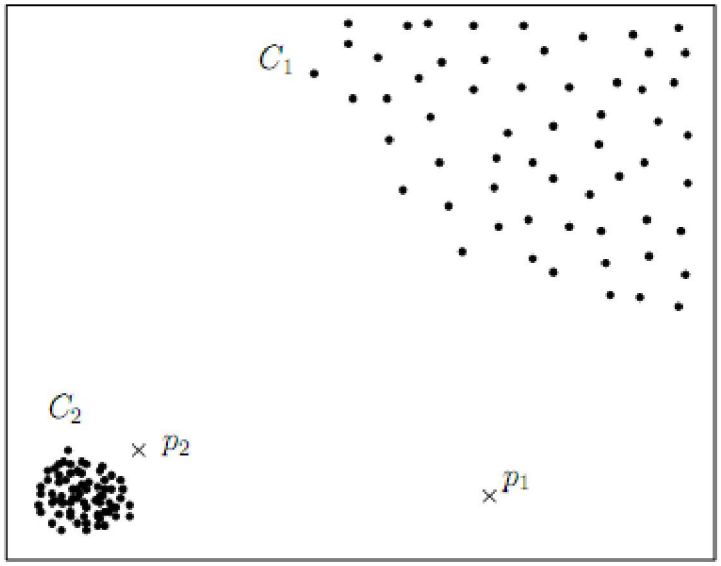

Figure 3.3 Nearest neighbor based approach [57]

The base assumption behind the working of these techniques is that normal data objects have many closely located neighbors while outliers are located in a comparatively low dense region which is normally far from normal regions. Detecting outliers using nearest neighbor approach is depicted in Figure 3.3.

Techniques of outlier detection based on nearest neighbor comprise of two steps, the first step is to compute neighborhood of each data object using a distance or a similarity measure (defined between two data instances) and then in the analysis of neighborhood is done to decide whether a data object is normal or outlier.

3.6.4 Classification-Based Outlier Detection Techniques

Classification [58] is a very significant machine learning and data mining notion. The key objective of classification is to learn a set of labeled data instances in the first phase known as training and then to classify an unseen data instance into one of the created classes which is the second phase known as testing.

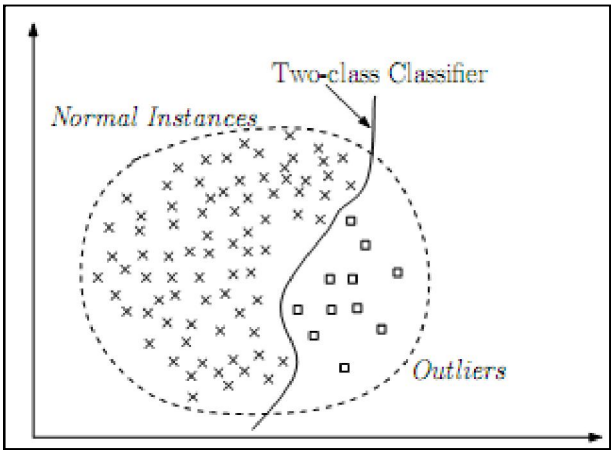

Figure 3.4 A two class classification based approach for outlier detection [58]

Classification based outlier detection techniques also run in the similar two phase fashion, using the same two classes of 'normal' and 'outlier'. A classification model is built in the training phase using the available labeled training data. A test instance is then classified using the learnt model in the test phase. Two class classification based approach for outlier detection is shown in Figure 3.4

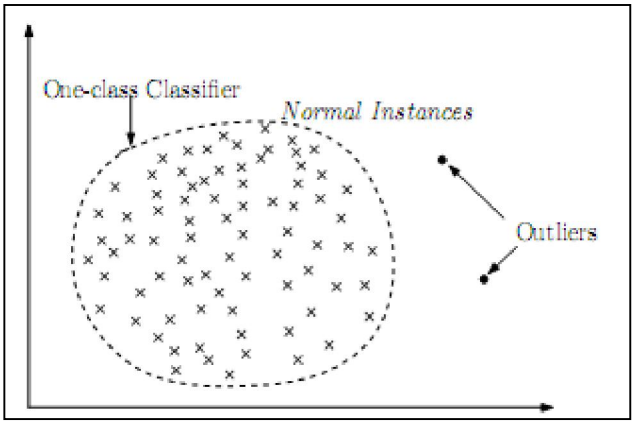

Figure 3.5 A one class classification based approach for outlier detection [58]

Conventional classification algorithms work on the basis of assumption of the availability of labeled normal and outlier data for training. Any outlier detection technique which works on such assumption i.e. the availability of both normal and outlier data object labels for training can be handled using rare class classification algorithms [59], [60], [61] because of the reason that there are reasonably fewer outliers than the normal data objects. One class classification based approach for outlier detection is shown in Figure 3.5.

Such techniques belong to the category of supervised outlier detection techniques. Sometimes, only outlier labels are available for training. Basically such techniques gain knowledge of the signatures of known outlier patterns and then generalize them to identify outliers in test data objects. Both Conventional and one class classification work on the assumption of the availability of labeled training data. The limitation of such techniques is the unavailability of labeled training data.

Such techniques also have another drawback of the uncertainty about how to classify an unseen data object. Because of such uncertainty a normal object can be classified as an outlier and an outlier as a normal instance leading to false positive and false negative accordingly.

Decision trees [58] is also a method that can be used to detect outliers in categorical data and so can identify errors and unexpected entries in data sets. While using decision trees for outlier detection, no prior knowledge of the data is required in contrast to many statistical methods that need parameters or distribution models derived from the data set. Decision trees have simple class boundaries as compared to the complex class boundaries generated by statistical methods.

Another outlier detection technique is rule-based systems [62], similar to decision trees because of the reason that both test a series of conditions; normally known as antecedents before a conclusion which is known as class is produce. Actually rules can be generated directly from the paths in the decision tree.

Rule-based systems are more flexible and incremental than decision trees as new rules may be added or rules amended without disturbing existing rules while a decision tree may require the generation of a complete new tree. The rule-based system may be a classifier which learns classification rules from both normal and outlier training data or can be a recogniser trained on normal data only and learning rules in order to detect the changes that identify anomalous activity.

3.6.5 Information Theory Based Techniques

Techniques based on information theory [63] perform analysis of the information present in a data set applying various information theoretic measures like entropy. These techniques are based on the idea that normal data object are regular with respect to certain information theoretic measure while outliers notably change the information contained in the data because of their

amazing nature. So these Techniques detect data objects that produce an irregularity in the data set, where the criterion for regularity is a specific information theoretic measure.

3.7 Anomaly Detection for Terrorist Event Prediction

Krebs, a very well known name wrote an article "Uncloaking terrorist networks" and added a section "Prevention or Prosecution" in his very interesting article. According to him, existing social network analysis (SNA) is applied more successfully to the prosecution not the prevention, of criminal activities [64]. After the September 11 attacks, a big question arose and is still unanswered, why wasn't this attack predicted and prevented? Why intelligence agencies lack the capability to uncover such covert plots ant stop them before they are executed?

This lack of prevention capability which is of course the proactive approach is the second motivational factor for this research. If something can be done before hand, maybe damage can be minimized and lives can be saved.

Anomaly detection has been applied in various domains as discussed earlier in this chapter but on the best of our knowledge, has not been applied for prediction of terrorist events. Following sections cover the proposed idea of applying outlier detection for detecting potential terrorist events.

The basic idea of anomaly detector is to detect anomalous activities of terrorist groups monitoring routine activities over a time line and predict event whenever an outlier occurs.
Suppose a detected terrorist group consisting of nine members. Suppose we have seven different databases which are integrated and which record the logs of activities of these members. An activity can be any action in which one actor performs an action on other actor or actors. For example, if a node 'A' sends an SMS to node 'B', it will be recorded as activity in the SMS log database, actors are nodes 'A' and 'B' and the weight of activity is one because of single SMS. Similarly if a node 'A' sends an email to multiple nodes, the activity will be recorded in the email database as an activity and the weight of this activity will be equal to the total number of emails i.e. the total number of nodes receiving that email, say for example if the email was sent to five node, the weight of overall activity will be equal to five. With these entire activities, time stamp will also be logged in the database in order to model activities on the time line. At any instance of time, the overall weight of a terrorist group is proposed to be equal to sum of all the

activities done of any member at that time. For example if in our case we are considering SMS, Telephonic conversation, Email, Bank transfer of an extra ordinary amount change of location the aggregate weight of activities on any time instance will be equal to sum of number of all SMS sent at that time + sum of number of all emails exchanged at that time + sum of number of telephonic conversations + sum of number of bank transfers made + sum of number of change of locations. Mathematical representation of same is shown in **Equation 3:1**.

$$w(t) = \sum_1^n s_i + \sum_1^n TC_i + \sum_1^n BT_i + \sum_1^n E_i + \sum_1^n L_i \qquad \textbf{Equation 3:1}$$

So after taking aggregate sum of numbers of all the activities done at a time instance, the activities will be modeled over a time line. The next step is the continuous monitoring the time line in order to detect any outliers which can be an indication of a possible threat.

The working of proposed Anomaly Detector consists of following steps:

1. **Data Logging:**

   All the activities of detected terrorist groups will be logged in the corresponding data sets. Following are the glimpse of the sample data bases used:

SMS database:

SenderNodeID	RecieverNodeID	TimeStamp	Text
1	2	1	Xyz
2	3	1	Xyz
2	4	1	Xyz

Email database:

SenderNodeID	RecieverNodeID	TimeStamp	Email
2	5	1	Some text
2	3	1	Some text

2	4	1	Some text

Telephonic conversation database:

CallerNodeID	CalleeNodeID	TimeStamp	Duration
5	6	1	0:30:19
7	8	1	0:5:1
8	10	1	0:3:9

Bank transfer database

TansferringNodeID	TransfereeNodeID	TimeStamp	Amount
5	10	1	USD 10,000
10	11	1	USD 5,000
10	12	1	USD 7,000

Travel record database

NodeID	Source	Destination	Time Stamp
12	sss	ddd	1
13	fff	ddd	1
14	Hhh	ddd	1

The data shown in all the above data bases depicts the activates performed at one time stamp, that one time stamp can be any time interval, may be an hour from 12:00 hours to 13:00 hours.

2. Integration

Data from all the sources will be integrated into one central data warehouse that will be used for central integrated information retrieval purpose.

3. Aggregation

Aggregated summary of all the activities done on different time intervals will be monitored on a time series. The input data will be fetched from the central data warehouse. In the data shown in the above tables, the magnitude of activities generated by different nodes of a terrorist group on time stamp 1 is equal to 15 because on this time, 3 SMS have been transferred, 3 emails have been transported, 3 bank transfers have been made, 3 telephonic conversations have been done and 3 nodes have changed their locations, so the aggregate sum is equal to 15.

The decision that this 15 is a normal or an outlier data object will be made with the help of outlier detection which will be discussed in the next section. However this is worth mentioning it here that if 15 is a normal values, which is representing that during one hour (we have taken time stamp equal to one hour) this terrorist group makes 3 SMS, 3 telephonic calls, 3 bank transfers, 3 email and 3 travels or may near these values. This normal value shows that the group is in passive mode. Passive mode can be the preparation or planning phase of a terrorist group during which they may be preparing for an attack or may be making future strategy but if this value is an outlier, i.e. if the number of activities performed during one hour is abnormal, the group may be going to carry out a terrorist attack so the law enforcement agencies should immediately eliminate the leaders that Group Leader Detection component of the proposed model has already been pointed out.

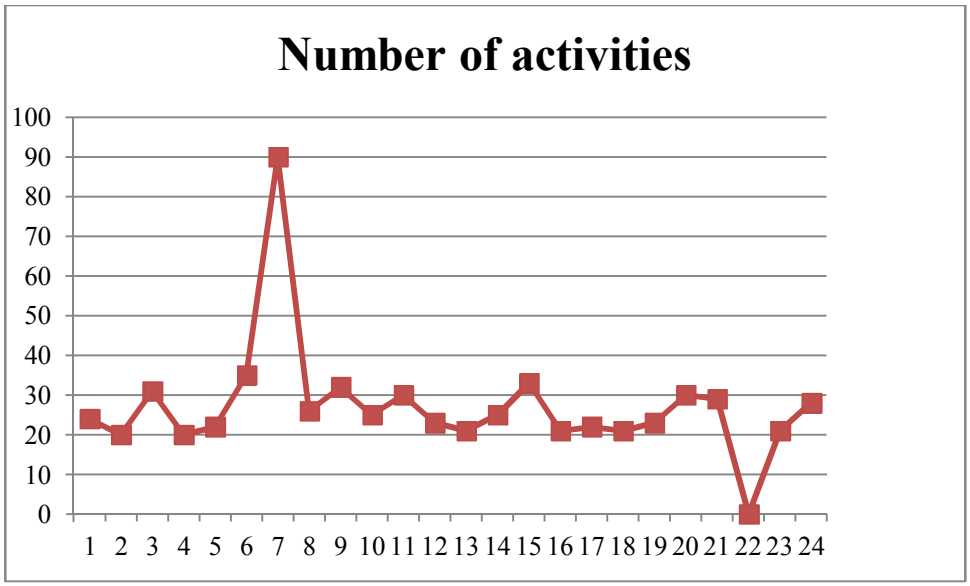

Figure 3.6 Activity monitory over timeline

In the chart shown in Figure 3.6 Activity monitory over timeline, x axis shows the time stamps while magnitude of activities done by members of a group are shown along y axis. Apparently the activities done at time interval 7 and 22 can be possible outliers indicating an indication of a terrorist attack but may be a false alarm.

Outlier detection methodology for the proposed model is discussed in the next section:

3.7.1 Outlier Detection

Before moving onto the details of outlier detection in the proposed framework, a brief discussion about outlier detection methodologies used in the proposed model is presented in the following sub sections.

3.7.2 Nearest Neighbor based outlier detection

Nearest neighbor based outlier detection has the key advantage of being purely data driven i.e. best for unsupervised detection.

As stated earlier; in Nearest Neighbor based outlier detection the base assumption is that normal objects have data objects have many closely located neighbors while outliers are located in a comparatively low dense region which is normally far from normal regions.

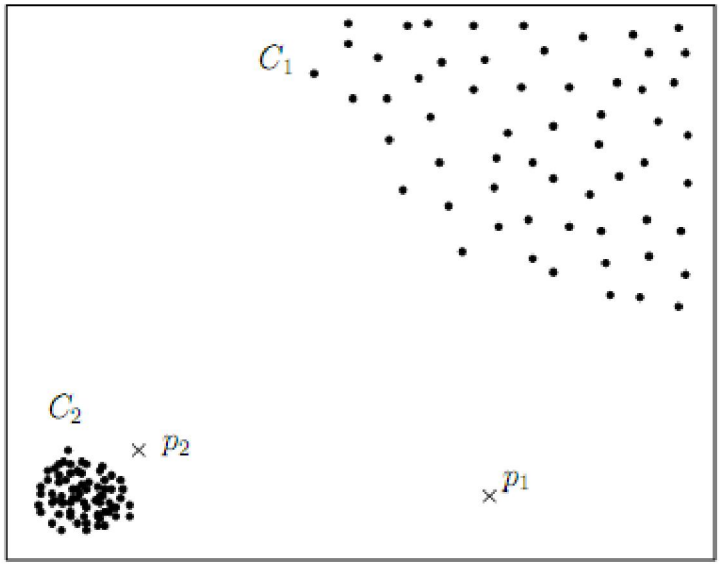

Figure 3.7 Nearest neighbor based approach

As Figure 3.7 indicates, C1 and C2 are two data clusters in which all the data objects are closely located indicating that all are normal data instances while points P1 and P2 are located in rare regions clearly depicting that they are outlier instances.

Outlier detection techniques based on nearest neighbor comprise of two steps, the first step is to compute neighborhood of each data instance (using a distance or a similarity measure defined between two data objects) and then in the second step the analysis of neighborhood is made to decide whether a data object is normal or outlier.

Outlier detection techniques that fall under category of nearest neighbor; operate using a distance or similarity measure which is defined between two data objects. There are different ways of computing distance or similarity between two data objects. Choice depends on the nature of data: generally:

- Euclidean distance is a mostly used for continuous data attributes but other measures can also be used [65]
- Generally a simple matching coefficient is used for data objects having categorical attributes. Some complex distance measures are also defined in [66] [67].

- Normally distance or similarity is calculated for each attribute and then combined for multivariate data i.e. data with multiple attributes.

Broadly nearest neighbor based techniques can be divided into two categories based on how they calculate the outlier score. The first category is "Distance to K^{th} Nearest Neighbor based". In these techniques, distance of a data object to k^{th} neighbor is calculated and is used as its outlier score. The second category is "Relative Density Based". In techniques belonging to this category relative density of each data object is computed to get its outlier score. Choice depends on nature of data or any other priority. Distance based outlier score calculation is used in the proposed model.

From neighborhood perspective there are three well known definitions of outliers:

1. The data objects in a dataset which have fewer than k neighbors where a neighbor is a data object that is within a distant R [68] [69].
2. Data objects are the n objects presenting the highest distance values to their respective k^{th} nearest neighbor [69].
3. Outliers are the n data objects in a data set that present the highest average distance to their respective k nearest neighbors [70].

The basic form of a K nearest neighbor outlier detection is known as simple nested loops (SNL) algorithm which has worst case complexity $O(n^2)$. The algorithm is quite basic so is not given here however a variant of the basis SNL algorithm known as canonical distance based algorithm is given below [71].

Procedure: Search for Outliers
Inputs: k, number of neighbours considered; n, number of outliers to be identified; D, the set of points
Outputs: O, the outlier result set
Let: Nearest(o, S, k) returns k elements from S that are the nearest to o
Let: Maxdist(o, S) returns the maximum distance between o and an element in set S
Let: TopOutlier(S, n) returns the top n outliers in S based on the distance to their k^{th} nearest neighbor.
Begin

```
1: O ← ∅ {Make the outlier result set empty}
2: D^k_min ← 0 {Reset the pruning threshold}
3: for each object o in D do
4:    Neighbours(o) ← ∅ {Make the neighbors's set from o
      empty}
5:    D^k(o) ← 0 {Reset the k nearest neighbor distance}
6:    {Searching for neighbours of object o}
7:    for each object v in D, where v ≠ o do
8:       Neighbours(o) = Nearest(o, Neighbours(o)∪ v, k)
9:       D^k(o) = Maxdist(o, Neighbours(o))
10:      if |Neighbours(o)| = k and D^k_min > D^k(o) then
11:         break {Discard this object as outlier, ANNS rule}
12:      end if
13:   end for
14:   O = TopOutliers(O ∪ o, n)
15:   if | O | = n then
16:      D^k_min = min(D^k(o) for all o in O)
17:   end if
18: end for
```

End

One of the most significant pruning rules in outlier detection was defined by Ramaswamy et al [69] referred as "Approximate Nearest Neighbor Search, ANNS" as can be seen in the above mentioned algorithm. ANNS focuses on just the top n outliers. ANNS rule states that we may disregard an object p as a candidate outlier if, while computing its $D^k(p)$, $D^k(p) < D^k_{min}$.

Although the algorithm has same quadratic worst case complexity but it does offer significant potential for optimization [71]. Some variations of nearest neighbor based outlier detection are presented below which have been used in the proposed framework.

3.7.3 K-NN Global Anomaly Score

This is the simplest form of nearest neighbor based outlier detection in which the outlier score is simply the average of the distance to the nearest neighbors. The value of k can be adjusted. The algorithm was first presented by Ramaswamy in [69]. The outlier score is calculated according to the measure type selected. The higher the outlier the more anomalous the instance is.

3.7.4 Local Outlier Factor

The LOF anomaly detection calculates the anomaly score according to the local outlier factor algorithm proposed by Breunig [72]. LOF is one of the earliest local density based approaches proposed. There are several steps in the calculation of the LOF. The initial step involves getting the nearest neighbors set. The definition of the k-distance employed is the one proposed in the original paper in order to handle duplicates. The definition states that the k-distance (p) has at least k neighbors with distinct spatial coordinates that have a distance less than or equal it and at most k-1 of such neighbors with distance strictly less than it. The reach ability distance (reach-dist(p,o)) is the maximum of the distance between point p and o and the k-distance(o). The local reach ability is the inverse of the average reach ability distance over the nearest neighborhood set. Finally the LOF is calculated as the average of the ratio of the local reach ability density over the neighborhood set. The values of the LOF oscillate with the change in the size of the neighborhood. Thus a range is defined for the size of the neighborhood. The maximum LOF over that range is taken as the final LOF score. A normal instance has an outlier value of approximately 1, while outliers have values greater than 1.

3.7.5 Connectivity Based Outlier Factor

The COF is a modification of LOF algorithm [72] proposed in order to handle outliers deviating from low density patterns. The definition of the k-distance used is the same as the one proposed by Breunig [72] to handle duplicates. The normal instances will have an outlier score of approximately 1, while outliers have a value greater than 1.

3.7.6 Histogram Based Statistical Outlier Score

In this method, a separate uni variate histogram is calculated for every column in the given data set. There are two modes, one with a static and one with a dynamic bandwidth. In the static mode, every bin has the same bin width equally distributed over the value range. In the dynamic mode, the bin width can vary, but you can specify a minimum number of examples contained in a bin. The parameter number of bins sets the total number of bins used for either mode. The bin width / minimum number values per bin are then calculated automatically. In the dynamic mode, it is possible that there are less bins then specified if some bins contain more than the minimum number of values. The default values for the number of bins are the square root of the number of total examples (number of bins set to -1). To compute the outlier score, the histograms are normalized to one in height first. Then, the score is inverted, so that anomalies have a high score and normal examples have a low score. It is also possible to apply a logarithmic scaling to the score in order to avoid floating point precision errors under certain circumstances. Furthermore, a ranked mode can be used for scoring. Here, the score is the sum of the ranks of an example among all (ordered) histograms instead of using the bin height.

3.7.7 Proposed Anomaly Detection Model

The above mentioned techniques are used in the proposed model to compute the outlier score in the given data set and the instances whose outlier score is remarkably high are marked is the points of potential threat. An ensemble of all the above discussed technique is also used in order to use advantages of all of them. Average of scores computed by all of the above mentioned technique is used to classify a data instance in normal or outlier object. On timeline, if an instance is detected to be outlier, the terrorist group under observation is classified to be in active mode while a normal instance on time line indicates a passive mode of that group.

3.8 Experimentation and Results

As discussed earlier, outlier detection is used to detect active modes of the terrorist groups which may be indicating the possible points of potential threats. Any outlier in the data under observation is considered to be the active mode which may be a point of attack. Due to unavailability of real terrorist's data, a similar real dataset is used for experimentation and

validation of the proposed. Details of software used, datasets, experiments and results are discussed in the following sub sections of this section.

3.8.1 Rapid Miner

Rapid miner is a software tool which provide platform to implement machine learning algorithms and applications [73].

3.8.2 Cyber Attackers Dataset

This dataset consists of real data log that has been created by IT department of our institute during detection and capturing of a hackers group who were intruding in our institute's management information system. Every time someone logged to the system was checked online and the login attempt was labeled as suspicious or safe login. The dataset consists of features usersid, hourof login, dayof login, IP Address of the node from where log attempt in was made and label representing a safe or a suspicious login attempt. The network involved in hacking is shown in Figure 3.8. A view of dataset is shown in Figure 3.9.The justification for using this real dataset instead of terrorist data is that the cyber attackers work just like a terrorist network. They coordinate in the same fashion as terrorists do. Terrorists coordinate among themselves to carry out terrorist attacks while cyber attackers coordinate to launch cyber attacks. Just like terrorist groups, cyber attacker groups have been seen in the same two modes, i.e. active and passive. Outlier detection has been applied on the discussed datasets to relate outliers with active modes. The dataset used is labeled i.e. it already has tags showing active or passive mode. A suspicious login is an active while a normal login is taken as passive mode.

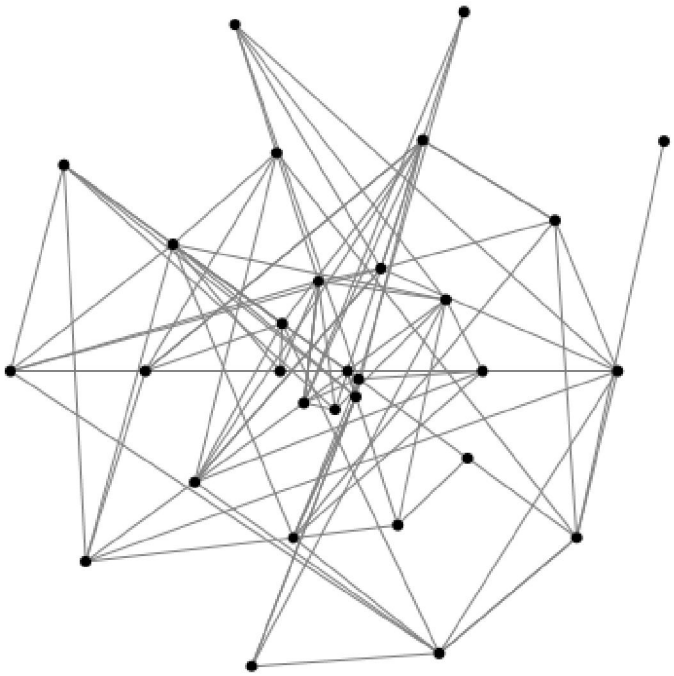

Figure 3.8: Cyber Attackers Network

UserID	Hour of Day	IP	Suspicious	Day
8	23	2	0	5
8	11	2	0	5
8	12	2	0	5
68	12	2	0	5
68	10	2	0	5
68	11	2	0	5
68	13	2	0	5
68	19	2	0	5
68	13	2	0	5
68	9	2	0	5
137	20	2	0	5
137	19	2	0	5
137	11	2	0	5
137	12	2	0	5
137	12	2	0	5
137	10	2	0	5
137	20	2	0	5
137	11	2	0	5
137	8	2	0	5
149	21	5	0	5
149	22	15	0	5
61	16	19	0	5
149	10	19	0	5
149	10	19	0	5

Figure 3.9: Glimpse of login log of dataset under discussion

The accuracy of the proposed model is taken as percentage of number of correctly identified modes out of total instances. All the measures discussed in the previous chapter were applied and used and separate accuracies were calculated. Table 3.1, Table 3.2, Table 3.3 and Table 3.4 show the accuracies of outlier detection methods keeping 15, 10, 5 and 2.5% top outlier score instances as outliers summary of results.

Method	Correctly identified instances	Total instances	Accuracy
k-NN Global Anomaly Score	9213	11040	83%
Local Outlier Factor	9341	11040	84%
Connectivity Based Outlier Factor	9341	11040	84%
Histogram Based Outlier Score	9067	11040	82%
Ensemble	9141	11040	82.7%

Table 3.1 Accuracies of outlier detection methods keeping top 15% instances as outliers

	Correctly identified instances	Total instances	Accuracy
k-NN Global Anomaly Score	9657	11040	87%
Local Outlier Factor	9857	11040	89%
Connectivity Based Outlier Factor	9857	11040	89%
Histogram Based Outlier Score	9529	11040	86%
Ensemble	9629	11040	87%

Table 3.2 Accuracies of outlier detection methods keeping top 10% instances as outliers

Method	Correctly identified instances	Total instances	Accuracy
k-NN Global Anomaly Score	10047	11040	91%
Local Outlier Factor	10275	11040	93%
Connectivity Based Outlier Factor	10276	11040	93%
Histogram Based Outlier Score	9995	11040	90%
Ensemble	10073	11040	91%

Table 3.3 Accuracies of outlier detection methods keeping top 5% instances as outliers

Method	Correctly identified instances	Total instances	Accuracy
k-NN Global Anomaly Score	10277	11040	93%
Local Outlier Factor	10403	11040	94%
Connectivity Based Outlier Factor	10403	11040	94%
Histogram Based Outlier Score	10239	11040	92.7%
Ensemble	10275	11040	93%

Table 3.4 Accuracies of outlier detection methods keeping top 2.5% instances as outliers

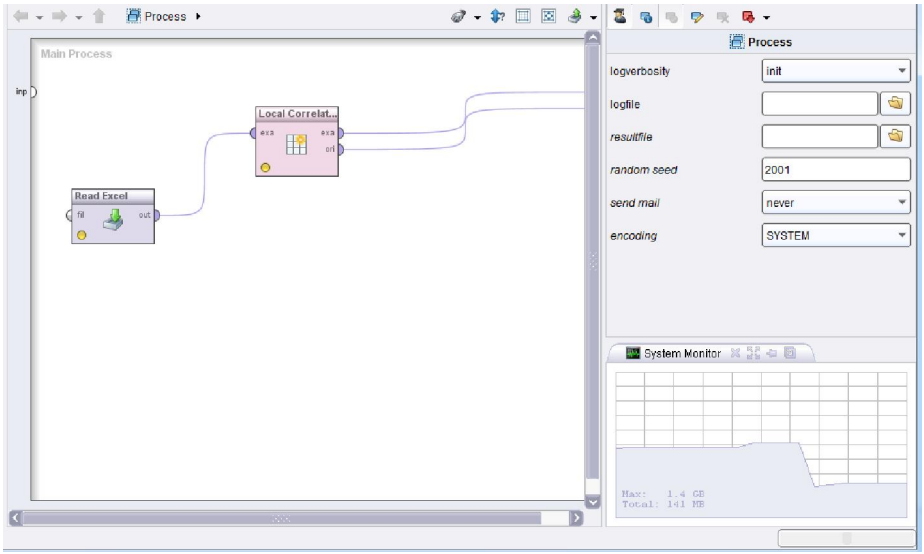

Figure 3.10 Outlier detection process in Rapid Miner

Figure 3.10 shows a view of implementation of outlier detection process in rapid miner.

Chapter 4 Classification and Prediction

Because of exponential growth in the size of databases across the globe, it is practically impossible to mine knowledge hidden in the patterns of the data stored without automatic methods of extracting information. Different algorithms have been created since last decades in order to process and mine the hidden patterns in the databases and convert them to useful knowledge. There are different methodologies which exist in the literature to which these knowledge extraction algorithms belong like: classification, association rules, clustering etc.

In this chapter, the basic concepts about data classification are presented. A novel hybrid classifier for key player detection problem given in chapter 2 has been proposed. Implementation, experimentation and results of the same are also presented. Also a novel hybrid classifier for prediction of terrorist group responsible for any terrorist incident is also proposed and presented in this chapter. Experiments, results and comparisons are also given.

4.1 Data Classification

Classification includes predicting an outcome based on a given input. A part of the given data set whose values which are to be predicted for unseen data are already known is used as the training set. A classification algorithm tries to find out relationship between the attribute whose value is to be predicted for unseen data (normally known as label) and other attributes. On the basis of that relationship, confidence for all possible values is calculated for label attribute of an unseen data instance and the label (class, prediction) which gets highest confidence value is assigned to that instance.

In the example shown in Table 1.1, a training set containing four attributes is given out of which the last attribute i.e. Heart Problem is the label or class attribute for which a classification algorithm will make a prediction. A classification algorithm will take this training set as an input and will process it to learn the relationship among Heart Problem and the other attributes. Whenever an unseen data instance is given to the algorithm, as some unseen instances are shown in the prediction set, the algorithm will calculate confidence for each possible label (Yes and No in the given case) on the basis of values for other given attributes and the relationship learnt from the training data. Label with highest confidence will be assigned to the unseen data instance.

Training set

Age	Heart rate	Blood pressure	Heart problem
65	78	150/70	Yes
37	83	112/76	No
71	67	108/65	No

Prediction set

Age	Heart rate	Blood pressure	Heart problem
43	98	147/89	?
65	58	106/63	?
84	77	150/65	?

Table 4.1 Training and Prediction Sets

Table 4.1 shows an example of training and prediction sets. Normally a classification algorithm uses prediction rules to represent knowledge [74]. These rules are stored and represented as IF-THEN rules. The IF part consists of a conjunction of conditions and the THEN part gives a certain prediction of value for the label of a data instance whose other attribute values satisfy that IF part. For example for the data shown in table 4.1, a rule predicting a value is given below:

```
IF (Age=65 AND Heart rate>70) OR (Age>60 AND Blood pressure>140/70)
THEN Heart problem=yes
```

The if part of this rule is satisfied for the first row of the prediction set, so "Yes" is the value predicted for Heart Problem attribute of this prediction instance as declared by THEN part of this rule.

Because of this pattern of working i.e. training and then prediction based on that training, classification is also known as supervised learning. The process of supervision consists of using training data (the labeled data) to train a classifier (a classification algorithm, technique or model) and then assigning labels to unseen unlabeled data. A classifier being trained and assigning labels is shown in Figure 4.1.

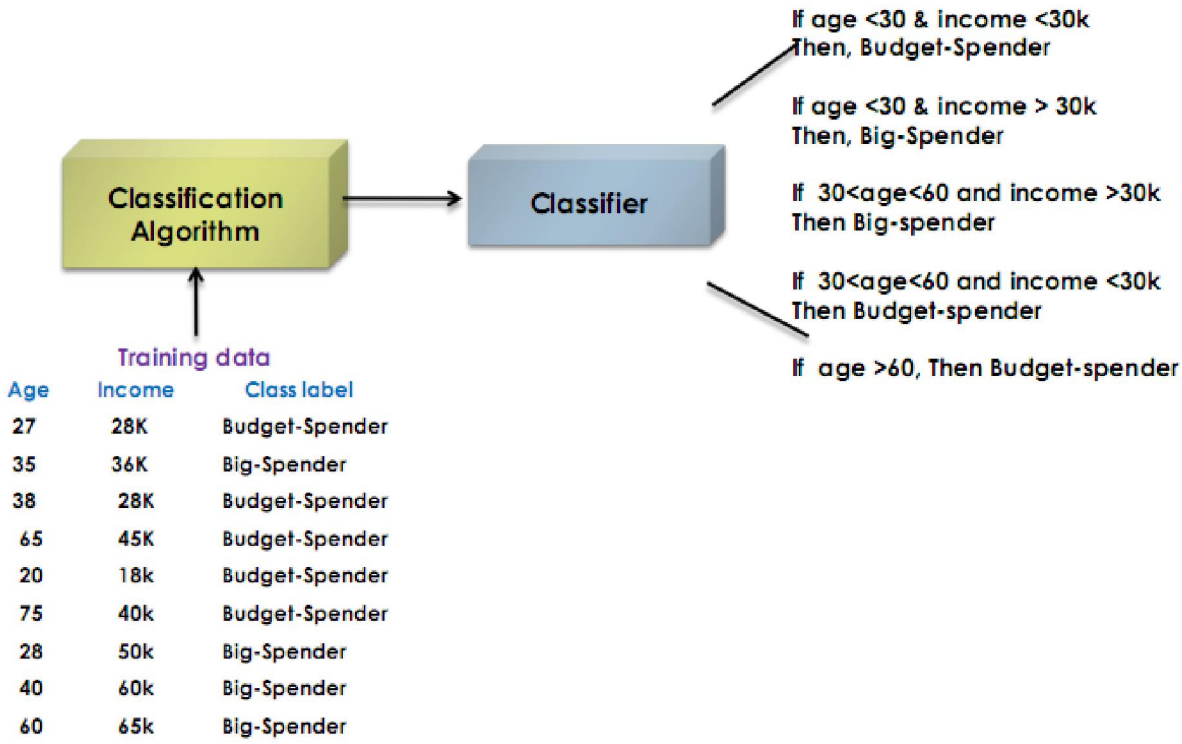

Figure 4.1 Training Classifier and Assigning Labels [75]

As discussed earlier, the first step in any classification is the model construction which is the learning or training phase. The tuples which are already labeled and their class is known because of those labels are used to construct the classification model. A classification model is generally represented by Classification Rules (As shown earlier), Decision Trees or Mathematical Formulae.

Before actually applying the classifier on actual unseen data, normally the trained classifier is also tested and evaluated. This step is generally termed as the testing phase. During test phase, the conceived model is applied on some tuples whose classes are already known just to check the validity and accuracy of the trained classifier. Test data is same as the training data in its structure. The predicted values for class labels by the trained classifier are compared to the actual values of the test instances in order to measure the accuracy. Figure 4.2 shows a classifier under testing.

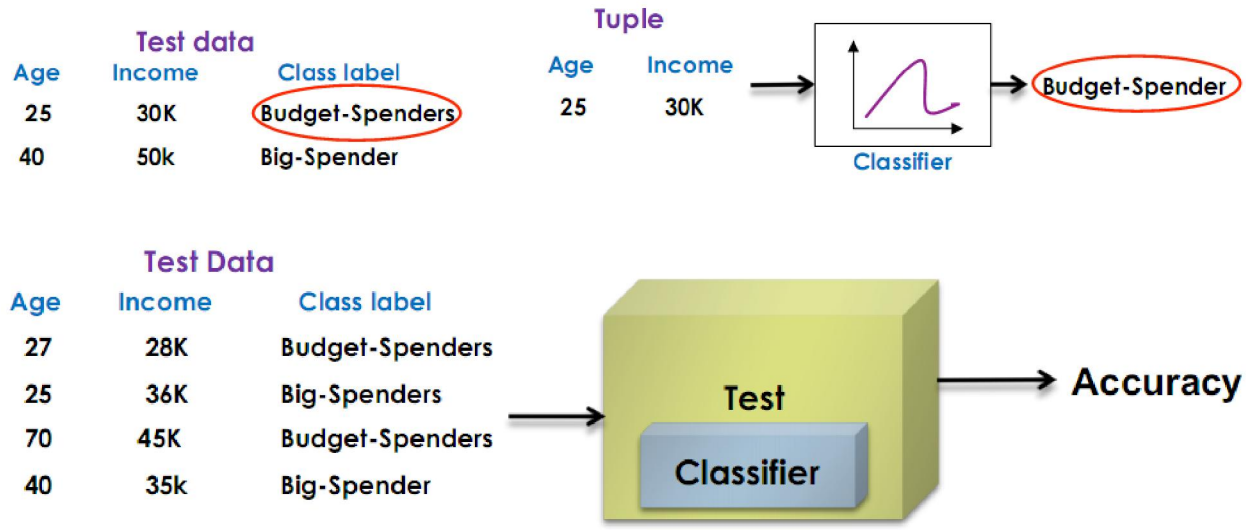

Figure 4.2 Testing a Classifier using Test Data [75]

Accuracy rate is the percentage of test data instances which are correctly classified by the classifier. An important point here in testing is the separation of test data from training data. If different data is not used for training and testing, chances of over-fitting are there. If the trained classifier achieves the preset threshold of accuracy then it may be used to predict labels on the real unseen data. Classifying unseen data by a classifier is shown in Figure 4.3.

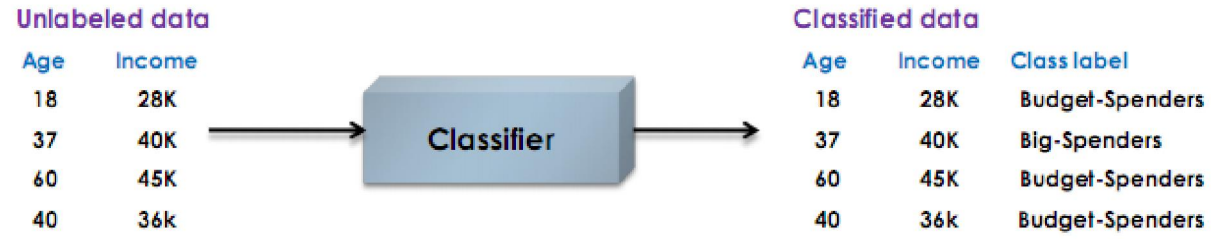

Figure 4.3 Using the trained and testes classifier on real unseen data [75]

Some issues related to classification and prediction are discussed in the following section: -

4.2 Issues of Classification and Prediction

Some important issues related to using Classification for a Prediction involve dealing with the following:

4.2.1 Data Cleansing

Datasets often contain some noisy and silent data. By noisy we mean some outlier data induced in the data set due to any measuring, collecting or human error. Silent data means data having missing values for some of its attribute. So before actually using the data from a dataset for training or testing, this very important issue needs to be dealt with. Outlier detection to remove noisy and filling missing values with most probable values for that attribute is normally applied as a preprocessing step before actually using the data to construct the classifier. This important issue must be resolved as a preprocessing step otherwise wrong results can be faced.

4.2.2 Relevance Analysis

Relevance Analysis [76] is also an important issue related to the data preparation step in data classification also known as feature selection. By relevance Analysis we mean to deal with the attributes contained in the dataset. There can be different attributes or feature which may be of no use to us because of their redundancy or irrelevance to the type of analysis we intend. So this important issue should also be dealt before actually giving the data set to the algorithm.

4.2.3 Data Transformation and Reduction

This issue is related to the type of values of different involved features of the target dataset. This is also a very important issue as classifier may not be able to deal non numeric or continuous data (depends on the classifier). So to deal this issue, data may be discretized, generalized or attribute values may be normalized.

4.2.4 Classifier Evaluation

To evaluate a classifier, following measures are normally used:

4.2.5 Accuracy

Accuracy of a classifier means how correctly a classifier predicts or labels unseen data. Accuracy can be calculated using one or more testing datasets that are independent of the training data.

4.2.6 Speed

Speed means the computational cost incurred in creating and using the classifier.

4.2.7 Robustness

Robustness means the ability of a classifier to work even if noisy data or data with missing values is given an input.

4.2.8 Scalability

Scalability refers to the ability of a classifier to deal with immensely large amount of data from very large databases.

4.2.9 Interpretability

This refers to the level of understanding or insight provided by a classifier.

4.2.10 Approaches

Two approaches are normally found in literature in order to solve the classification and prediction problems [77].

4.2.11 Train and Test

This approach simply is to partition the data set D into two disjoint training and testing subsets. Firstly training the classifier using the training subset and then measuring performance of the classifier using the test subset partitioned.

4.2.12 M-Fold Cross Validation

This approach consists of partitioning the dataset D into m fragments. Then using different classifiers and training them using different fragment from m. Then testing each trained classifier using test set and other fragments from m.

4.2.13 Classification Accuracy

As stated earlier, accuracy is a measure to determine how fit a classifier is to be used for a specific problem. Mathematically:

Let Trg be the training data which is a subset of dataset D, 'Te' be the test dataset which also is a subset of D and C(o) denotes the original class of object o. Then classification accuracy of classifier C on Te is given is **Equation 4:1**.

$$Accuracy_{Te}(Classifier) = \frac{|\{o \in Te | Classifier(o) = C(o)\}|}{|Te|} \qquad \textbf{Equation 4:1}$$

The classification error of the same is given in **Equation 4:2**:

$$Error_{Te}(Classifier) = \frac{|\{o \in Te | Classifier(o) \neq C(o)\}|}{|Te|} \qquad \textbf{Equation 4:2}$$

4.2.14 Confusion Matrix

Let $c_1 \in C$ be the target positive class and union of all other classes be the contrasting negative class. If we compare the predicted and actual class labels, we can distinguish the results into four different cases as shown in Table 4.2.

	Predicted as positive	Predicted as negative
Actually positive	True Positive (TP)	False Negative (FN)
Actually Negative	False Positive (FP)	True Negative (TN)

Table 4.2 Confusion Matrix

4.2.15 Precision and Recall

These two measures are used with respect to the target class while using a classifier. Precision is given mathematically in **Equation 4:3**.

$$Precision(Classifier) = \frac{TP}{TP+FP} \qquad \textbf{Equation 4:3}$$

And the recall is given in **Equation 4:4**.

$$Recall(Classifier) = \frac{TP}{TP+FN} \qquad \textbf{Equation 4:4}$$

There is a tradeof between these two measures. These two measures are related with each other as shown in **Equation 4:5**

$$F - Measure(Classifier) = \frac{2.Precision(Classifier).Recall(Classifier)}{Precision(Classifier)+Recall(Classifier)} \qquad \textbf{Equation 4:5}$$

Using these measures, accuracy can also be viewed mathematically as shown in **Equation 4:6**.

$$Accuracy = \frac{\text{True Positives+True Negatives}}{\text{True Positives+False Positives+True Negatives+False Negatives}} \qquad \textbf{Equation 4:6}$$

4.3 Classification Approaches

Various well known classification approaches from literature are presented in this section.

4.3.1 Classification by Decision Tree Induction

In decision tree induction, decision trees are learnt from class labeled training instances. A decision tree resembles a flow chart like tree. In decision trees, every non-leaf node represents a test on an attribute, each branch is representation of an outcome of a test and each leaf node contains a class label. A typical form of decision tree is shown in Figure 4.4

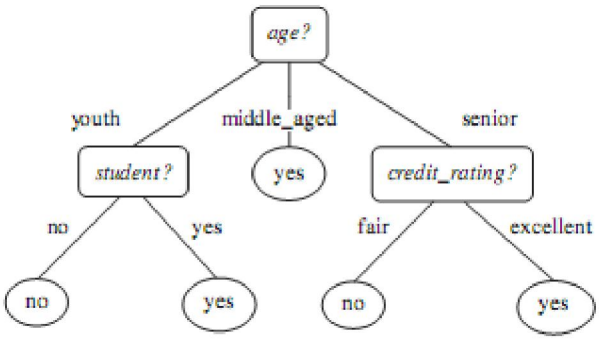

Figure 4.4 A decision tree for the concept buy_computer, indicating whether a customer at All electronics is likely to purchase a computer. Each internal (nonleaf) node represents a test on an attribute. Each leaf node represents a class (either buys computer=yes or buys computer=no) [78]

Decision trees as can be seen can be used very conveniently for classification. For any data instance for which class label is unknown, the attribute values are tested against the decision. A

path is marked from the root to the leaf where the decision of the label of that instance lies. Decision trees can be very easily interpreted into classification rules.

The reason of popularity of decision trees is that any domain knowledge or parameter setting is not required to construct a decision tree so decision trees are appropriate for exploratory knowledge discovery. High dimensional data can be handled by decision trees. The process of training and classification of decision is normally simple and fast. Also generally decision trees are found to be having a very good accuracy. Decision tree for classification has been widely used in various areas including medicine, manufacturing, financial analysis, astronomy, molecular biology and so on.

A very famous decision tree algorithm known as ID3 (Iterative Dichotomiser) was developed by Ross Quinlan, a very active machine learning researcher during the early 1970s. Same researcher presented C4.5, a successor aor ID3 which became a benchmark for all new supervised learning algorithms.

4.3.2 Bayesian Classification

Bayesian Classifiers are statistical classifiers. They are used to predict class membership probabilities, for instance, probability of a specific data instance of belonging to a specific class. Bayes' theorem is the foundation of these classifiers. These classifiers generally have been found highly accurate and speedy when applied to large data sets.

The Bayes theorem upon which these classifiers are based as mentioned earlier was named after Thomas Bayes who did early work in probability and decision theory during the 18th century. For every data instance X, P(H|X) is computed which is the posterior probability of H conditioned on X. P(H) is in contrast prior probability of H. Also P(X) is the probability of X. Bayesian Theorem is given in **Equation 4:7**

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \qquad \textbf{Equation 4:7}$$

Naïve Bayes Classification algorithm is a very famous Bayesian theorem based classification algorithm. The simple Naïve Bayes classifier works in the following way:

1. Let suppose D is a training set of data instances and their respective class labels. Each data instance is represented by an n dimensional attribute vector, $X=(x_1,x_2,...,x_n)$, representing n measurements made on data instances from n different attributes respectively. Where n different attributes are $A_1, A_2,...,A_n$.
2. Let suppose there are m number of classes i.e. $C_1,C_2,...C_m$. For any data instance X, the classifier will predict that X belongs to a class having highest posterior probability conditioned on X. Naïve Bayes classifier predicts that a specific data instance X belong to a class Ci if and only if $P(C_i|X) > P(C_j|X)$ for 1<=j<=m. j< >i.

   Therefore $P(C_i|X)$ is maximized. Any class C_i for which $P(C_i|X)$ is maximized is called posteriori hypothesis. Representation of same in terms of Bayes theorem can be seen in **Equation 4:8**

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)} \qquad \textbf{Equation 4:8}$$

3. In **Equation 4:8**, as denominator is constant for all the classes, so numerator needs to be maximized. In case if class's probabilities are unknown, it is assumed that probability of all classes is equally likely.

Bayesian classifiers results have been found comparable in some domains with other well known classifiers. Also in theory Bayesian classifiers have minimum error rate but it's not always true in practice.

4.3.3 Rule Based Classification:

In Rule Based Classification, the learnt classifying model is represented as a set of IF-THEN rules. Rules have been found very good for representing knowledge or information. A rule based classifier uses IF-THEN rules for classifying any data instance. The classifier itself gets train by IF-THEN rules which may be learnt from decision tree or through other rules. Generally rules are of the form **IF *condition* THEN *conclusion.*** An example can be seen in **Equation 4:9**

IF age = youth AND student = yes THEN buys_computer = yes **Equation 4:9**

The IF part of the rule in known as rule antecedent or precondition. The THEN part is known as the rule consequent. The rule antecedent part consists of one or more attribute tests which are

ANDed in case there are tests on more than one attribute while the consequent contains prediction about a specific instance of belonging to a specific class label. The rule given in **Equation 4:9** can also be written as shown in **Equation 4:10**.

$$(\text{age} = \text{youth}) \wedge (\text{student} = \text{yes}) \Rightarrow (\text{buys computer} = \text{yes}) \qquad \textbf{Equation 4:10}$$

Rules are evaluated by two parameters namely coverage and accuracy. For a given data instance X, from a labeled training data set D, let suppose n_{covers} be the number of data instances covered by a rule R; $n_{correct}$ be the number of data instances labeled correctly by the rule R and |D| be the number of total data instances in D. The coverage and accuracy of rule R can be defined as given in **Equation 4:11** and **Equation 4:12** respectively.

$$coverage(R) = \frac{n_{covers}}{|D|} \qquad \textbf{Equation 4:11}$$

$$accuracy(R) = \frac{n_{correct}}{n_{covers}} \qquad \textbf{Equation 4:12}$$

As it can be seen from the **Equation 4:11** and **Equation 4:12** that a rule's coverage is the percentage of data instances which are covered by that rule and accuracy shows the percentage of data instances which that rule classifies correctly.

Normally there are two ways of rule generation. The first as discussed earlier is the rule extraction from a decision tree and the other is rule induction using sequential coverage aalgorithm.

While extracting rules from decision tree, one rule is created for each path of the decision tree i.e. path from the root to every leaf node. Each criterion along the path is logically ANDed and the leaf node contains the prediction i.e. the THEN part of the rule.

Another way of rule creation is creating them directly from the training data using a sequential coverage algorithm. As rules are learnt sequentially so this algorithm got its name. This approach is the most widely used approach. Many sequential covering algorithms can be found in the literature. Some popular variants are AQ, CN2 and RIPPER. Generally the strategy is to learn one rule at a time. Each time when a rule is learnt, the data instances covered by the rule are removed from the data set and the procedure continues on the remaining data instances.

4.3.4 Classification by Back Propagation

Back propagation is basically a neural network learning algorithm. A neural network is a set of inter connected input/output units in which every connection has an associated weight. In the training phase, weights are adjusted by the network in order to predict the correct label for unseen data instances.

Neural networks are usually used in the applications where long training time doesn't matter because of their nature which is they need long training times. Neural networks are normally not directly understandable by human reading so they are criticized for this.

The advantages of neural network include their high tolerance for noisy data and also their capability of classifying patterns for data upon which they have not been trained. They are parallel in nature, so parallelization techniques may be used to speed up the computation process.

4.3.5 Support Vector Machines

A support vector machine (SVM) uses a non linear mapping to transform the original training data into a higher dimension and then in that higher dimension searches for linear optimal separating hyperplane which is used as a decision boundary separating data instances from one another. The hyper plane which has been mentioned is found by SVM using support vectors.

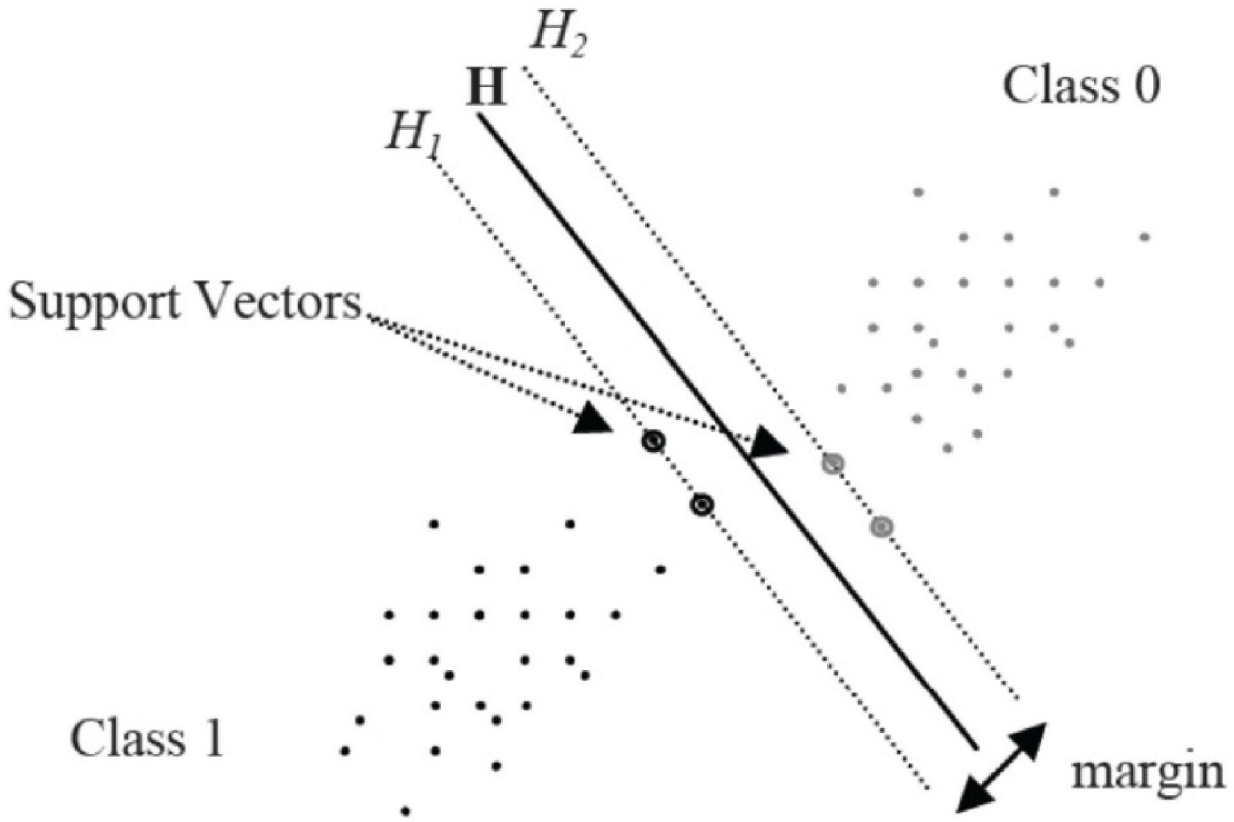

Figure 4.5 Generic Support Vector Machine working example [79]

An example of generic SVM can be seen in Figure 4.5.

4.3.6 Classification by Association Rule Analysis

Frequent patterns and their respective association rules also characterize interesting relationships between feature conditions and class labels. So these also have been used for classification. Association rules are used to represent set of features or attributes which have strong associations between their value pairs that frequently occur in a given dataset. Generally this method is used widely to make analysis of customers' purchase patterns. Such analysis can be helpful in decision making processes. Association rule set detection is based on frequent itemset mining.

$$age = youth \wedge credit = OK \Rightarrow buys\ computer = yes[support = 20\%, confidence = 93\%] \quad \textbf{Equation 4:13}$$

A typical association rule can be seen in **Equation 4:13** where '∧' represents logical AND.

4.3.7 Lazy Learners (Learning from neighbors)

Lazy learner unlike previously discussed methods waits until the last minute before doing any model construction in order to classfiy a given test data instance. When a lazy learner is given a training instance, it simply stores it and waits until it is given with a test data instance. In order to classify a test instance it uses its similarity to the stored training instances. Unlike other methods it does less work on the training intances, rather it works more while making classification or prediction. K nearest neighbor classifier is the most common example of lazy learners. K nearest neighbor compares the test intance with its K neighbors (K is the number of neighbors considered) and then based on similarity makes a decision about the classification. Eucledian distance is commonly used as distance measure.

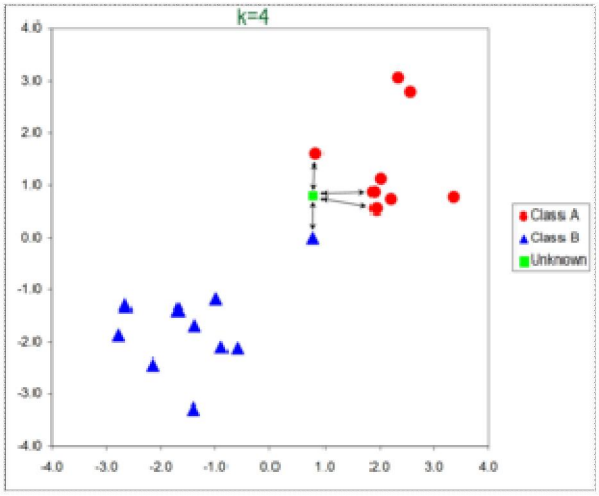

Figure 4.6 glimpse of working of K Nearest Neighbor where K=4. [80]

Figure 4.6 shows an example of K nearest neighbor working when applied on an un seen data instance and k has been taken equal to 4.

4.4 Proposed Model for Key Player Detection using Hybrid Classifier

Key player detection is an important step while analyzing covert networks. The proposed framework for key player detection consists of centrality measures for each node followed by hybrid classier for accurate detection of key players.

4.4.1 Data Preprocessing

The first step before applying proposed key player detection is data pre processing. The purpose of this step is to clean the data in order to facilitate further steps. Data preprocessing consists of

- Redundant feature removal
- Removal of duplicate entries
- Handling missing values

First step uses two rank sum tests i.e. Wilcoxon Rank-Sum and Ansari-Bradley Tests. Wilcoxon Rank-Sum test is a non-parametric test of the null hypothesis that two populations are the same against an alternative hypothesis that the two distributions differ only with respect to the median. It has higher efficiency on non-normal distributions such as a mixture of normal distributions [81]. Ansari-Bradley test compares two independent samples which come from the same distribution against the alternative that they come from same distributions having the same median and shape but different variances [82].

Preprocessing step also checks for duplicate entries and removes all such entries to avoid redundancy. The last step in preprocessing is to handle missing values in the data. The preprocessing technique identifies the missing feature values and then they are replaced by the mean value for that feature. This procedure is performed for those attributes where values are missing in less than 50% of the instances. If the number of instances with missing values is more than or equal to 50%, the particular attribute is rejected and not used further. Figure 4.7 shows the flow diagram of data preprocessing to handle any kind of data redundancies.

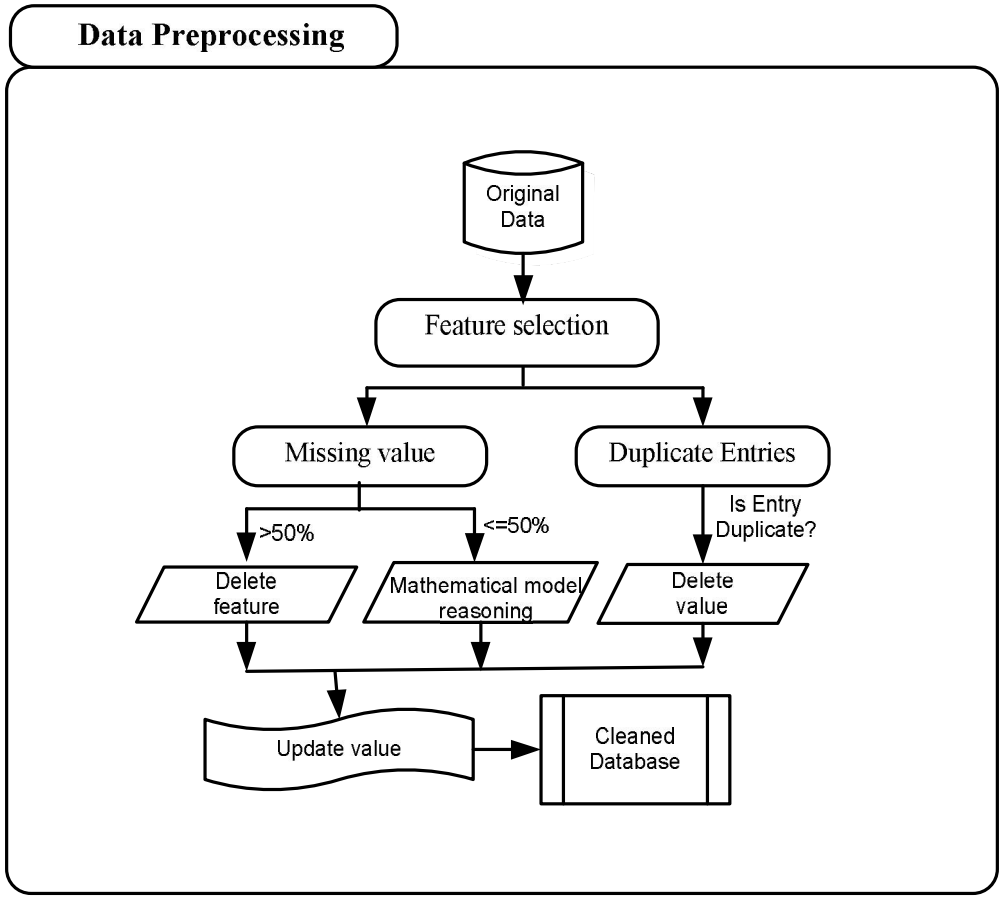

Figure 4.7 Flow diagram for handling data redundancy

4.4.2 Key Player Detection

Key player detection is an important step while analyzing covert networks. The proposed framework for key player detection consists of centrality measures for each node followed by hybrid classier for accurate detection of key players.

The four centrality measures which we have included in our proposed model are degree centrality (DC), betweenness centrality (BC), closeness centrality (CC) and eigen vector centrality (EC). Figure 4.8 shows the flow diagram of proposed model for key player detection.

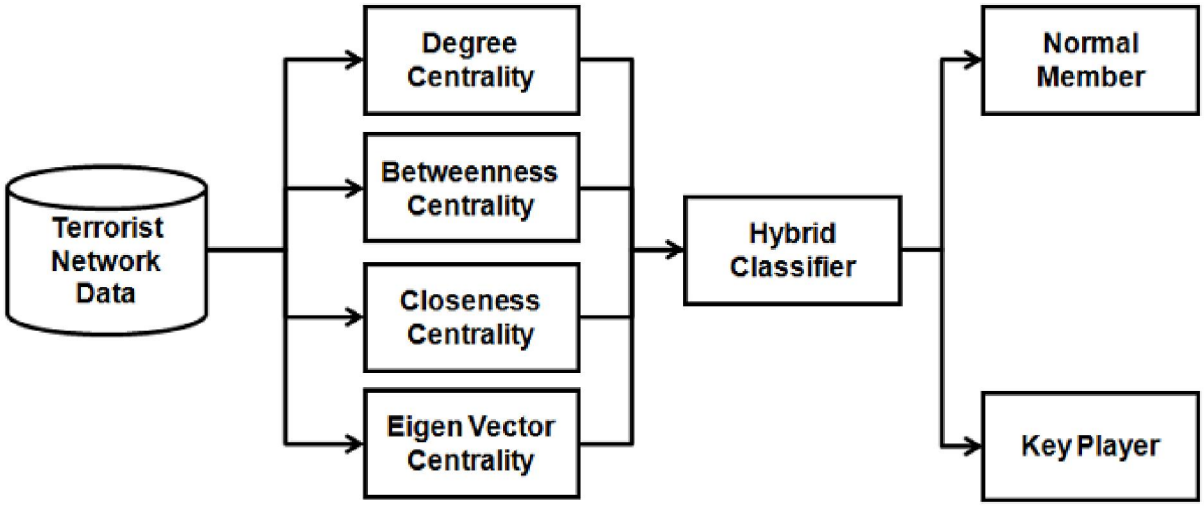

Figure 4.8 Proposed hybrid classifier

Key players normally appear as most central nodes in any network so they have significant values of centrality measures. If a covert network X contains k nodes, then the set representation for that network X is $X=\{v_1,v_2,...,v_k\}$. Here v_i represents i^{th} node in the given network. for an automated system to analyze each node as key player or normal node, a feature set is formed for each node. Each node is considered as sample for classification and represented by a feature vector containing four features i.e. for a sample node v from a network X, the feature vector is v={DC,BC,CC,EC}.

Once all nodes are represented by feature vectors, next phase is to classify them as key player or normal member. A new hybrid classifier as an ensemble of k-nearest neighbors (KNN), Gaussian mixture model (GMM) and support vector machine (SVM) is proposed here for accurate detection of key players. The purpose of using these three classifiers is to accurately model the distribution of data and to find accurate decision boundary by using the strengths of all three classifiers. KNN has simple implementation and gives good results when ever samples of same class exist as closest neighbors. GMM is famous due to the capability of accurately representing the data distribution and it caters for overlapping patterns where modeling of distribution gives a

good clue. SVM caters for the data which is well separable by a decision boundary and has good classification and rapid training phase.

4.4.3 k Nearest Neighbors (kNN)

k Nearest Neighbors (kNN) kNN is the most simplest and fundamental classifier used for supervised classification [83]. It is a kind of voting based classifier which finds k nearest samples from complete dataset based on some distance transform and assigns majority vote class to test sample. Let υ_i be a feature vector for i^{th} node with m features (f_{i1} , f_{i2} , f_{i3} ,.. , f_{im}), n be the total number of nodes (i = 1, 2, .., n) and m the total number of features (j = 1, 2,.., m). The Euclidean distance between node υ_i and υ_l where (l = 1, 2, .., n) is defined in **Equation 4:14**.

$$d(x_i, x_l) = \sqrt{(x_{i1} - x_{l1})^2 + (x_{i2} - x_{l2})^2 + \cdots + (x_{ip} - x_{lp})^2}.$$

Equation 4:14

Now depending upon the value of k, we choose closest k samples and assign the majority class to unknown node.

4.4.4 Gaussian Mixture Model (GMM)

To implement GMM, we use a two class Bayesian classifier using Gaussian functions [84]. Bayes decision rule is stated as [85] is shown in Equation 4:15

$$\begin{array}{llll} Choose & R_1 & if, p(\mathbf{v}|R_1)P(R_1) > p(\mathbf{v}|R_2)P(R_2) \\ otherwise & choose & R_2 \end{array}$$

Equation 4:15

where p(v|R_i) is the class conditional Probability Density Function (pdf) also known as likelihood and P (R_i) is the prior probability of class R_i which is calculated as the ratio of class R_i samples in the training set. The class conditional pdf of the feature vector for different classes is computed using multivariate Gaussian pdf [85] shown in Equation 4:16.

$$N(\mathbf{v}|\mu, \Sigma) = \frac{1}{(2\pi)^{\frac{m}{2}}|\Sigma|^2} exp[\frac{-1}{2}(\mathbf{v} - \mu)\Sigma^{-1}(\mathbf{v} - \mu)]$$

Equation 4:16

Where v and μ are feature vector containing m number of features and mean vector containing mean of each feature respectively. Σ is an m × m covariance matrix. In our case m= 4. We model the class conditional pdf's as linear combination of weighted Gaussian functions to represent the likelihood of a GMM using **Equation 4:17**.

$$p(\mathbf{v}|R_i) = \sum_{j=1}^{\kappa_i} N(\mathbf{v}|\mu_j, \Sigma_j)\omega_j$$

Equation 4:17

where K_i is the number of Gaussian mixtures used for Bayesian classification and $p(v|R_i)$ is a m- dimensional Gaussian distribution of weight ω_j and $R_i = \{R_1 , R_2 \}$ are the two classes used in proposed system.

The parameters for GMM are optimized using Expectation Maximization (EM) which is an iterative method and it chooses optimal parameters by finding the local maximum value of GMM distributions for training data. The EM starts with initial values of parameters (μ, Σ) and weight w for each Gaussian. In estimation step, EM computes the probability (PE) of each point for each Gaussian using **Equation 4:18**.

$$P_E(n, j) = \frac{w_j N(\upsilon_n|\mu_j, \Sigma_j)}{\sum_{i=1}^{\kappa} N(\upsilon_n|\mu_i, \Sigma_i)\omega_i}$$

Equation 4:18

Here PE (n, j) represents the probability that n^{th} candidate region υ_n is generated from j^{th} Gaussian. We do this for all K Gaussians and candidate regions. The second step is the maximization of likelihood by changing the parameters. The mean, Covariance matrix and weight for j^{th} Gaussian are updated using estimated probabilities and are given in **Equation 4:19**, **Equation 4:20** and **Equation 4:21** respectively.

$$\mu_j = \frac{1}{\xi_j} \sum_{n=1}^{N_{Total}} P_E(n, j)\upsilon_n$$

Equation 4:19

$$\Sigma_j = \frac{1}{\xi_j} \sum_{n=1}^{N_{Total}} P_E(n,j)(\upsilon_n - \mu_j)(\upsilon_n - \mu_j)^T$$

Equation 4:20

$$\omega_j = \frac{\xi_j}{N_{Total}}$$

Equation 4:21

Where $\xi_j = \sum_{n=1}^{N_{Total}} P_E(n,j)$ and N_{Total} are the total number of nodes.

5.1.1. Support Vector Machine (SVM)

SVM is used as third classifier in proposed framework for key player detection. The original algorithm of SVM separates different regions from each other with maximum margin by using a separating hyperplane if the classes are linearly separable. Due to close relevance of nodes, the proposed features make a nonlinear hyperplane for which SVM is applied along with kernel function based on radial basis function (RBF). To implement SVM along with RBF, we have applied least squares SVM using LS-SVM toolbox [86]. In LS-SVM, the multiclass solution is found by solving a system of linear equations instead of original quadratic programming.

5.1.2. Hybrid Classifier (HF)

For hybrid classifier, we combine kNN, GMM and SVM classifier using a weighted probabilistic ensemble. The classification of node υ using probabilistic classification prediction, based on measure of evidence from different classifiers, is performed as shown in **Equation 4:22**.

$$class(\upsilon) = \arg \max_{\forall class_i} \left(\sum_{i=1}^{c} a_k * P_{\mathbf{C_k}}(y = class_i | \upsilon) \right)$$

Equation 4:22

where $P_{\mathbf{C_k}}(y = class_i|\upsilon)$ is the probability of class i given a sample node using classifier k and a_k is the weight associated to the probabilistic prediction of class C_k . Figure 4.9 shows the proposed ensemble framework for hybrid classifier.

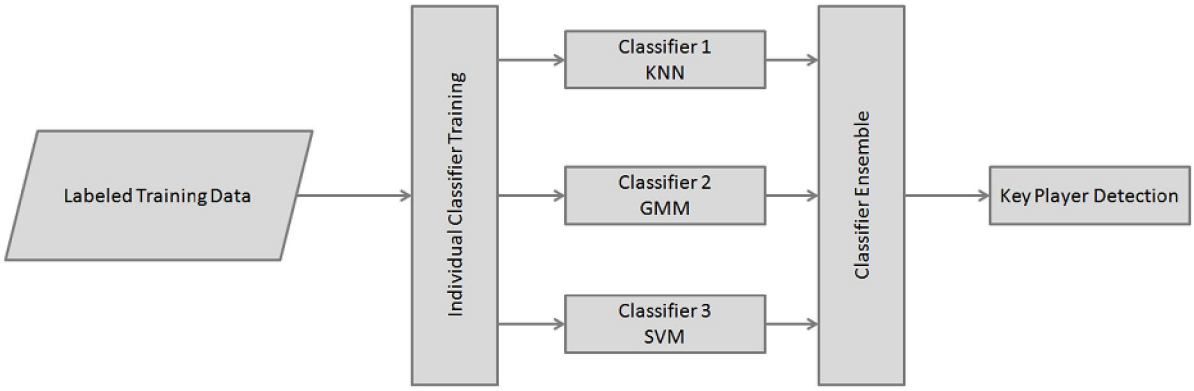

Figure 4.9 Proposed framework for hybrid classifier

These weights are optimized using genetic algorithm. The modeling of proposed hybrid classifier consists of two phases. In first phase, the algorithm separates out all confusing samples from complete training data. Confusing samples are those samples for which all three classifiers (kNN, GMM and SVM) give different decision and only these samples are used to optimize the weights for each classifier. This selection of confusing samples reduces the time for genetic algorithm in finding optimized weights. The second phase applies genetic algorithm using classification accuracy as objective function and purpose is to find such weights which maximize this objective function. The initial population consists of 20 weight vectors in which 16 are generated randomly and remaining four are [1, 0, 0], [0, 1, 0], [0, 0, 1] and [0.33, 0.33, 0.33]. Last four weight vectors are added to give maximum and equal confidence to all classifiers. All weight vectors are normalized to have a sum equal to 1. New population is generated using top 10 weight vectors from initial population based on objective function and applying crossover and mutation on remaining vectors. This is done to keep track of best vectors. The iterations are performed until there is no improvement in classification accuracy given in eq. 14 for ten consecutive iterations or the algorithm reaches to maximum iteration which is set equal to 100.

4.4.5 Learning Optimized Weights using Genetic Algorithm

The proposed ensemble framework given in **Equation 4:22** consists of a feature vector $a_k = \{a_{kNN}, a_{GMM}, a_{SVM}\}$. These weights are optimized using genetic algorithm. The modeling of weights consists of two phases i.e. separation of confused samples and learning of optimized weights. In first phase, the algorithm separates out all confusing samples from complete training

data. Confusing samples are those samples for which all three classifiers (kNN, GMM and SVM) give different decisions and only these samples are used to optimize the weights for each classifier. This selection of confusing samples reduces the time for genetic algorithm in finding optimized weights. The second phase applies genetic algorithm for learning of optimal weights.
The parameters of genetic algorithm such as definition of population, size of population, rules for crossover and mutation and objective function are defined as:

- **Population:** Each chromosome consists of a weight vector of three members which are weights for each classifier. All weight vectors are normalized to have a sum equal to 1.
- **Population Size:** The initial population consists of 20 normalized weight vectors in which 16 are generated randomly and remaining four are [1; 0; 0], [0; 1; 0], [0; 0; 1] and [0:33; 0:33; 0:33]. Last four weight vectors are added to give maximum and equal confidence to all classifiers.
- **Crossover:** Single point uniform crossover is used during learning. Crossover point is after first weight element which means that two selected chromosomes interchanges their weights for GMM and SVM classifiers. The selection of chromosomes for crossover is performed based on objective function value. Worst 10 chromosomes out of population of 20 are selected for crossover.
- **Mutation:** The mutation probability of 0% is used for mutation which means that no change is made in off springs after crossover.
- **Objective Function:** The classification accuracy corresponding to a specific weight vector is taken as objective function as defined in **Equation 4:22** and we want to maximize this function.

The iterative learning is performed until there is no improvement in classification accuracy given in Equation 4:22 for ten consecutive iterations or the algorithm reaches to maximum iteration which is set equal to 100.

4.4.6 Experimentation and Results:

This section covers experimentation and results related to the second proposed model which is for key player detection using a hybrid classifier.

4.4.7 Material

Same software NodeXL which plugs in with Microsoft excel is used for testing of key player detection. For proper evaluation of proposed framework, we have used three case studies. The nodes in all three networks are labeled as key players and normal members. Table 4.3 shows the network specification for all three case studies on which the proposed system is evaluated.

Case study	Nodes	Edges	Key players
I	79	402	8
II	62	150	19
III	30	114	9

Table 4.3 Network Specifications

4.4.8 Case study 1

The first case study is taken from [87] titled as Noordin Muhammad network. This subset of the Noordin Top Terrorist Network were drawn primarily from "Terrorism in Indonesia: Noordin's Networks", a 2006 publication of the International Crisis Group. It includes relational data on the 79 individuals listed in Appendix C of that publication. The data were initially coded by Naval Postgraduate School students as part of the course "Tracking and Disrupting Dark Networks" under the direction of Professor Sean Everton, Co-Director of the CORE Lab, and Professor Nancy Roberts. CORE Lab Research Associate Dan Cunningham also reviewed and helped clean the data. Figure 4.10 shows network generated in NodeXL for Noordin's network.

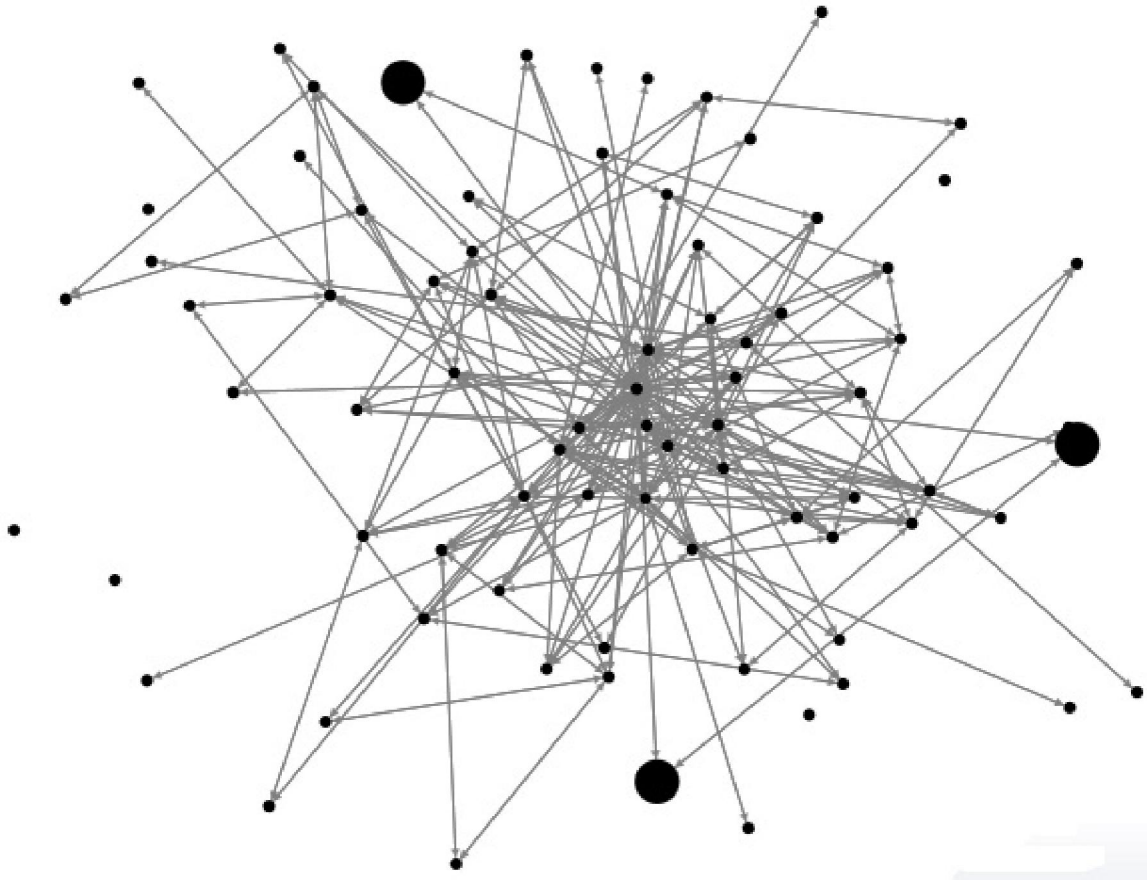

Figure 4.10 Noordin Muhammad Network

4.4.9 Case study 2

The dataset for second case study was first compiled by Valdis Kerbs [64] consisting the tragic 9-11 attackers network. The overall network consisted of 62 nodes and 150 edges containing all the attackers and their helpers who helped or coordinated in any way to organize the attacks. Muhammad Atta was the leader as confirmed by Ossama Bin Laden in a video tape [64]. The actual 19 hijackers who got crashed are labeled. They are considered important because they are the actual implementers of the attack. Figure 4.11 shows network generated in NodeXL for 9-11 attackers' network.

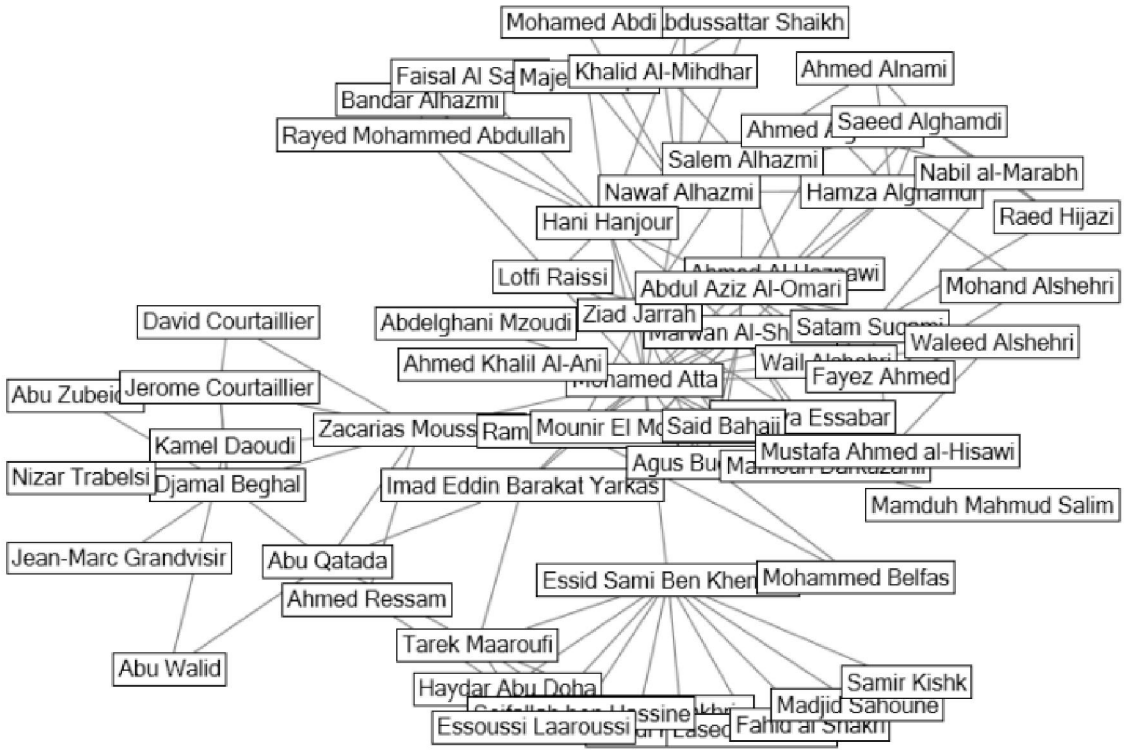

Figure 4.11 9/11 Network

4.4.10 Case study 3

This dataset consists of real data that has been created by IT department of our institute during detection and capturing of a hackers group who were intruding in our institute's management information system. The network present in the dataset was created when a hacker was traced on a complaint; all his connections were traced from his communication links and his In/Out Data log. The network consists of 30 nodes and 114 edges.

4.4.11 Results

The detailed quantitative and comparative analysis of proposed system is performed in this section. The performance of proposed system is measured using sensitivity (sen), specificity(spec), accuracy (acc) and area under receiver operating characteristics (ROC) curves (AUC) as figures of merit. Sensitivity is true positive rate and specificity is true negative rate. These parameters are calculated using **Equation 4:23**, **Equation 4:24** and **Equation 4:25** respectively.

$$Sensitivity = \frac{T_P}{(T_P + F_N)}$$ **Equation 4:23**

$$Specificity = \frac{T_N}{(T_N + F_P)}$$ **Equation 4:24**

$$Accuracy = \frac{(T_P + T_N)}{(T_P + T_N + F_P + F_N)}$$ **Equation 4:25**

• TP are true positives means number of key players which are identified correctly.

• TN are true negatives means number of normal members of network which are identified correctly.

• FP are false positives means number of normal members of network which are wrongly identified as key players.

• FN are false negatives means number of key players which are wrongly identified as normal members of network.

The modeling of classifiers has been done using randomly selected 70% of data as training and remaining 30% data as testing. The experiments are repeated 10 times and their average results are given. Table 4.4 shows the results of proposed framework for key player detection on all three networks given in case studies.

Case study	sen	Spec	acc	AUC
I	93.69	89.71	91.52	0.91
II	89.31	87.03	88.73	0.86
III	95.03	96.42	95.91	0.93

Table 4.4 Statistical performance evaluation of proposed framework for key player detection

The statistical analysis of proposed system is done with the help of ROC curves which are plots of sensitivity versus 1-specificity. This analysis is done for performance evaluation of proposed Hybrid Classifier (HC). Figure 4.12 shows the averaged ROC curves for all three case studies.

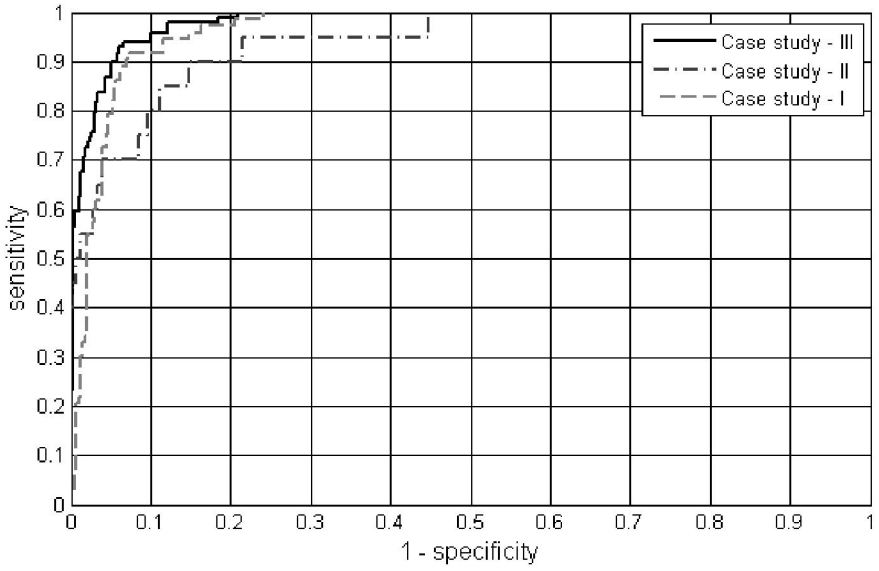

Figure 4.12 Averaged ROC curves for all three case studies

The proposed hybrid classifier is compared with individual KNN, GMM and SVM classifiers. The hybrid classifier is also compared with the existing well-established classifier ensemble methods such as ADABOOST [88], bagging [89] and random subspace methods (RSM) [90]. Table 4.5 shows the comparison of all these in terms of accuracy for key player detection.

Methods	case study-I	case study-II	case study-III
KNN	84.35	80.09	81.32
GMM	87.72	84.38	87.50
SVM	88.03	81.25	93.75
Adaboost	82.15	80.41	89.47
Bagging	86.78	81.57	90.72
RSM	86.13	82.61	87.03
Proposed Hybrid Classifier	91.52	88.52	95.91

Table 4.5 Comparison of hybrid classifier with existing ensemble methods

The proposed framework has been tested using three case studies and number of statistical measures. The results taken clearly prove the validity and correctness of proposed frame work. A new study from actual local event is taken and proposed system is tested on that as well. Along with proposed framework for key player detection, we have also tested a voting based method which takes care of all four centrality measures to detect the top k key players of the terrorist network. Here k is equal to the total number of key players present in a network which is under study. A node is considered as key player of three out of four centrality measures declared it as key player. This idea detected key players with accuracies of and 79%, 65%, 83.33% for all three case studies respectively. The proposed framework achieved accuracies of 91.52%, 88.73 and 95.91% for same case studies respectively. The results showed the validity of our framework and it can be used for detection key players for any suspicious network along with detection of any abnormal activity.

4.5 Proposed Ensemble Classifier for Terrorist Group Prediction

Terrorist Group Prediction ensemble classifier is proposed to be used for prediction of responsible terrorist group which may be responsible for any terrorist event. The proposed classifier is an ensemble of k-NN, Naïve Bayes and Decision Tree with gini index. The proposed model has been tested on a very well known publically available dataset known as Global Terrorism Database (GTD) which has data related to terrorist incidents occurred through 1970s till 2013. Details about dataset, experimentations and results are discussed in upcoming sections.

As discussed earlier, the proposed model for terrorist group prediction has three basic algorithms in its core. Two of them i.e. kNN and Naïve Bayes have also been used in the previous model i.e the key player detection model using hybrid classifier so they can be consulted from the previous section. Only the decision tree with Gini Index is presented in the next section.

4.5.1 Decision Tree with Gini Index

As discussed earlier, decision trees have got a very basic important role in data mining and pattern recognition. Commonly it has been observed that datasets contain very large number of features and mostly some of them do not contribute. So the number of those features can be reduced using Gini Index. Only informative attributes can produce rule which are compact

therefore only those attributes which have lower values for Gini Index will be used for rule generation as shown in **Equation 4:26**.

$$Gini(t) = 1 - \sum_{i=0}^{c-1} [p\left(\frac{i}{t}\right)]^2 \qquad \textbf{Equation 4:26}$$

Figure 4.13 shows glimpse of decision tree with gini index created for the GTD data.

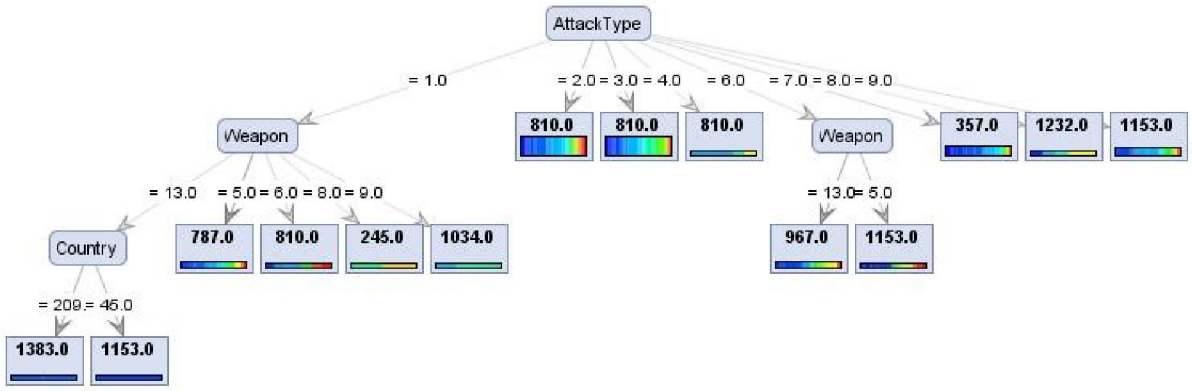

Figure 4.13 Decision Tree for TGP

4.5.2 The Proposed Ensemble method

The classifier ensemble method for Terrorist Group Prediction (TGP) method combines the three classifiers stated earlier. The proposed ensemble model is shown in Figure 4.14.

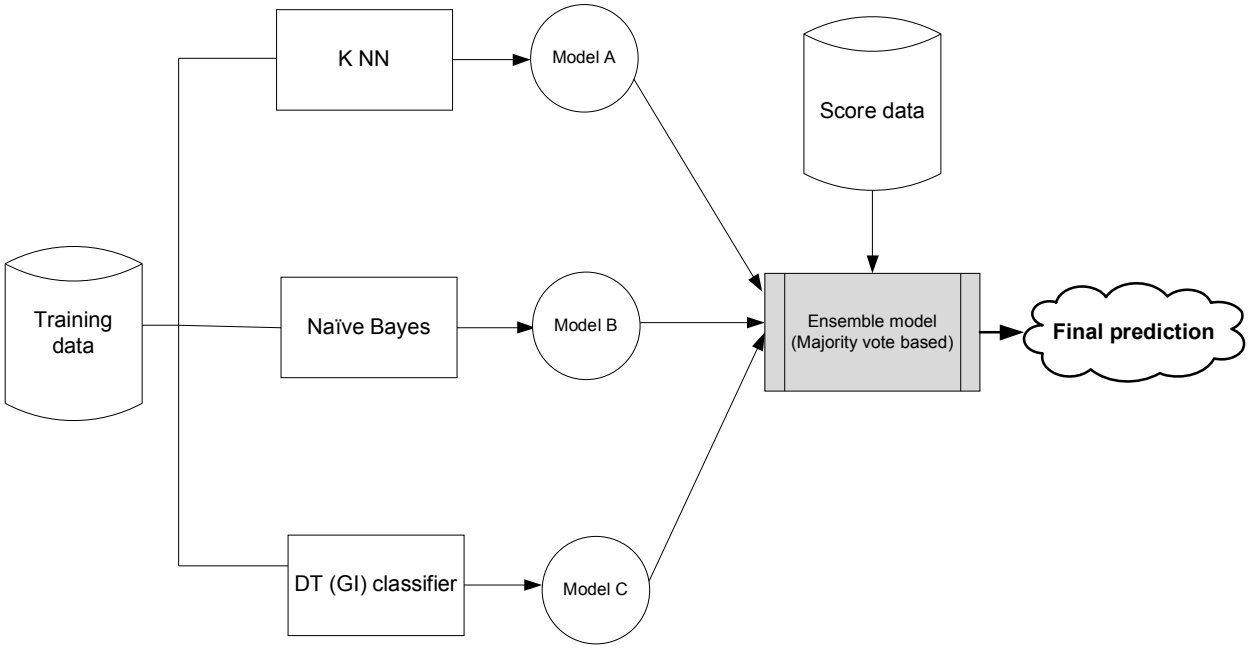

Figure 4.14 Proposed ensemble model

Ensemble approach consists of two steps. In the first step, individual members are generated and in the second step the individual member's output is combined to create a new model. All classifiers of the model are trained using the data set.

Let N denotes the number of classifiers represented by $C_1 \ldots\ldots C_N$. Let A is a set of classifiers i.e. $A=\{C_i: i=1;N)$ and let M denotes the number of output classes. The proposed ensemble method is defined as follows:

Find votes combination V for each classifier C_i optimizing to a function F(V). V's value is Boolean in type which represents the binary vote based ensemble. The size of V is N*M.

In boolean array, V(i,j) represents the decision that whether i^{th} classifier has voted for j^{th} class or not. V(i,j)=True/1 shows that i^{th} classifier has voted for j^{th} class; whereas V(i,j)=False/0 indicates that the i^{th} classifier has not voted for j^{th} class. Function F(V) represents the quality measure of classification technique for combined classifiers like sensitivity, specificity and accuracy.

Moreover, the main focus of proposed ensemble method is to create classifiers that differ on their decisions. In generalized way, these methods change the training process in order to create classifiers that result in different classifications.

Detailed architecture of the proposed model is shown is Figure 4.15 while flow chart of the proposed model is shown in Figure 4.16.

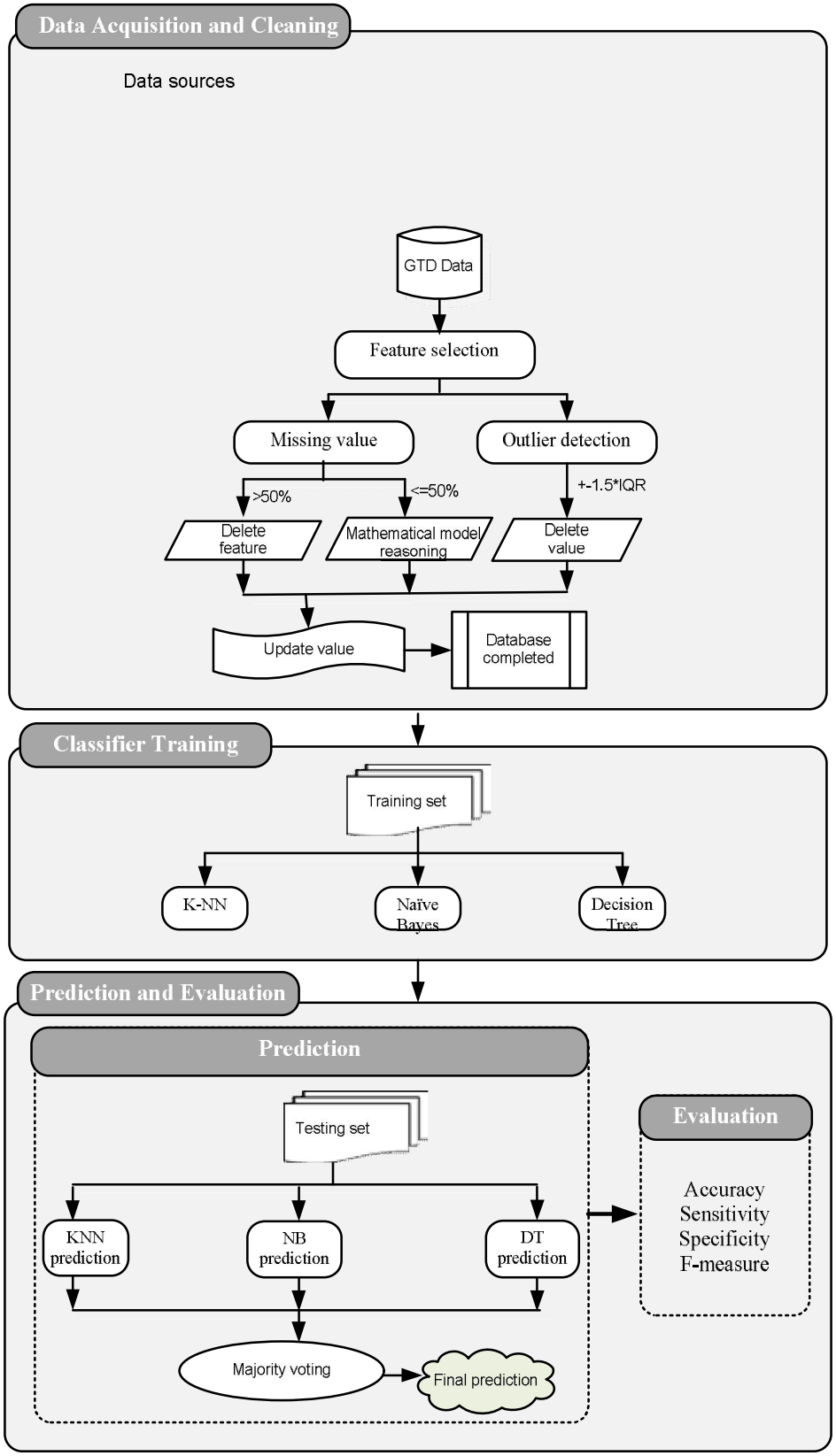

Figure 4.15 Detailed architecture of the proposed TGP model

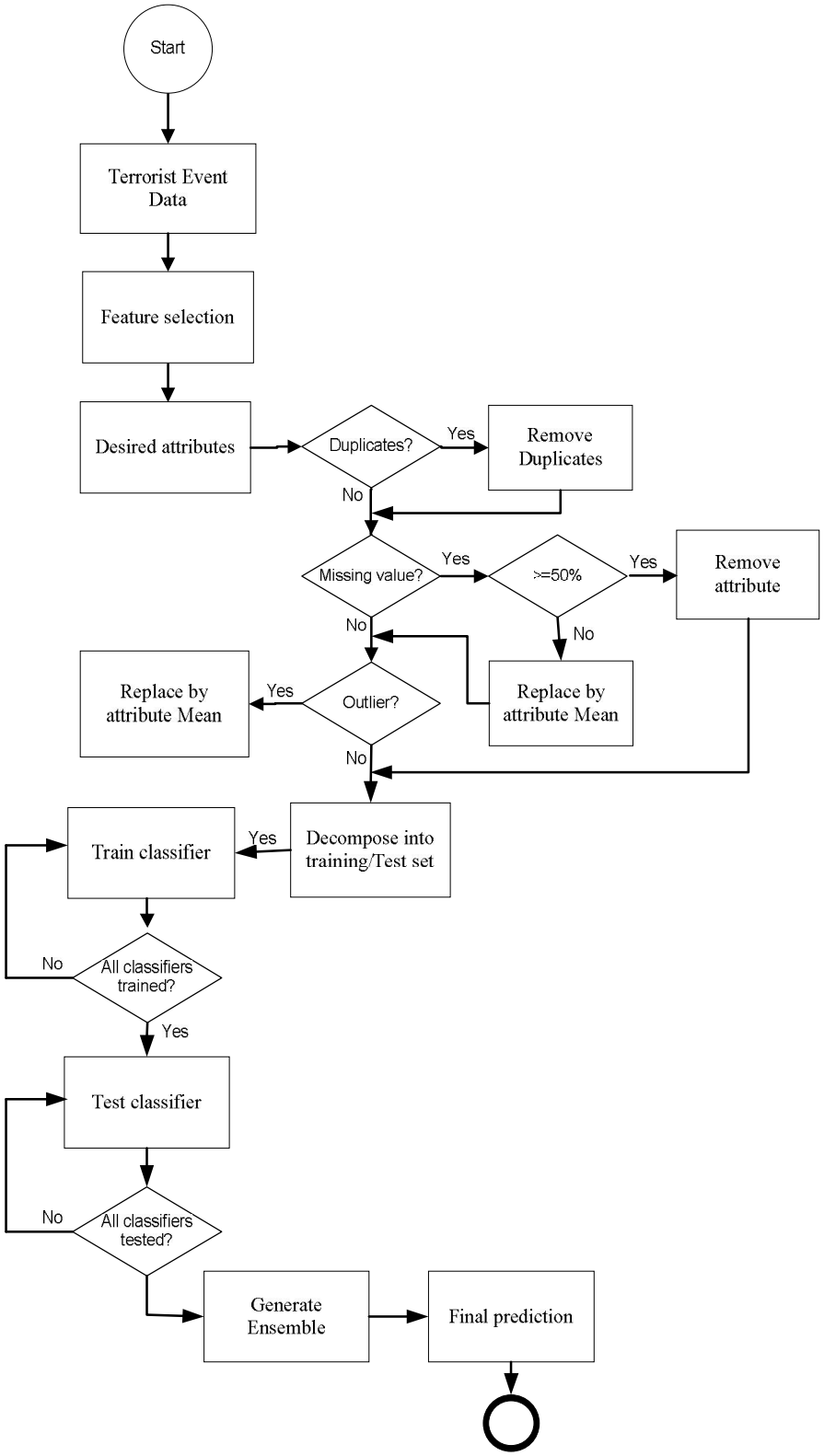

Figure 4.16 Flow chart of TGP

4.5.3 Experiments and Results

This section covers experimentation and results related to the third proposed model which is for terrorist group detection using a hybrid classifier.

4.5.4 Material

Rapid miner has been used to implement the proposed terrorist group detection model. The dataset used is the publicly available Global Terrorism Database which has been compiled by National Consortium for the Study of Terrorism and Responses to Terrorism (START): A Center of Excellence of the U.S. Department of Homeland Security University of Maryland. The dataset contains information on over 113,000 terrorist attacks. The dataset is claimed currently the most comprehensive unclassified data base on terrorist events in the world which includes information on more than 52,000 bombings, 14,400 assassinations, and 5,600 kidnappings since 1970. The data set also includes information on at least 45 variables for each case, with more recent incidents including information on more than 120 variables supervised by an advisory panel of 12 terrorism research experts. To collect this mass data, over 4,000,000 news articles and 25,000 news sources were reviewed. Figure 4.17 shows a screen shot of implementation of the proposed model using rapid miner.

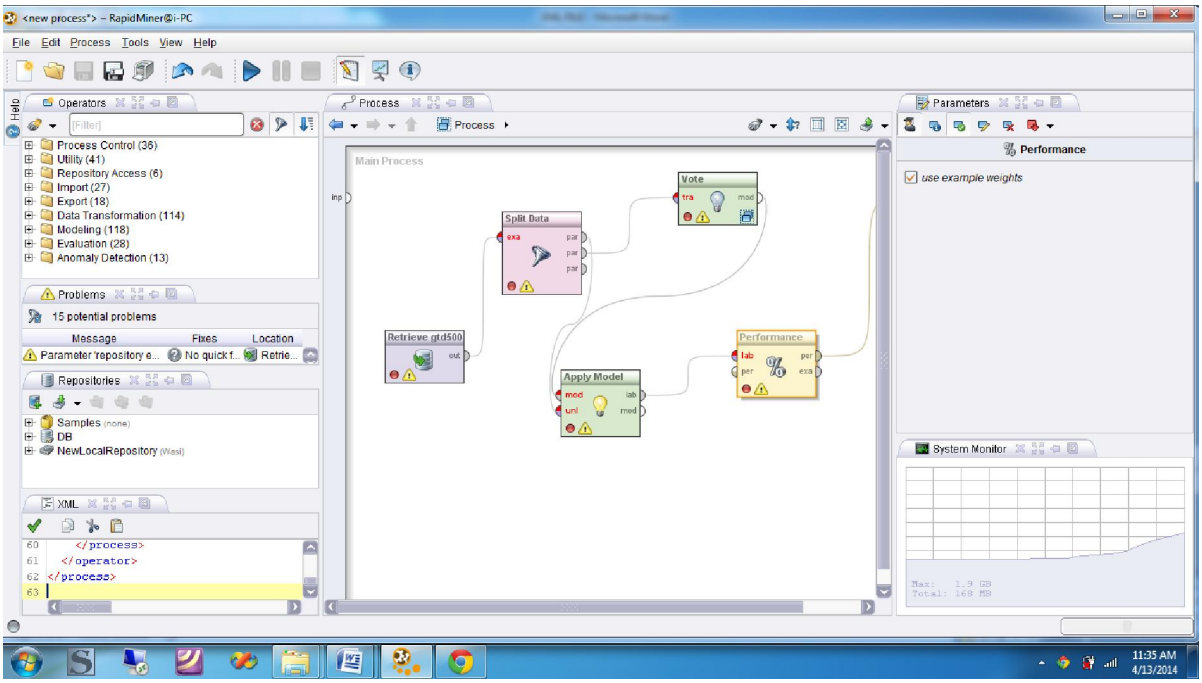

Figure 4.17 Rapid miner process for Terrorist Group Prediction

As mentioned in the previous chapter, in the proposed model, this first step is the data preprocessing. During this step the dataset was analyzed and instances with missing values were discarded. Out of available features, only seven relevant features are chosen which include Month, City, Country, Attack_Type, Group, Target and Weapon. Group is the label attribute. After implementation, some classes were found very small and ignorable compared to other classes. They were also ignored as their presence was affecting the results too badly.

The performance of proposed system is measured using sensitivity (sen), specificity(spec), accuracy (acc) and area under receiver operating characteristics (ROC) curves (AUC) as figures of merit. Sensitivity is true positive rate and specificity is true negative rate. These parameters are calculated using **Equation 4:27**, **Equation 4:28** and **Equation 4:29** respectively.

$$Sensitivity = \frac{T_P}{(T_P + F_N)}$$ **Equation 4:27**

$$Specificity = \frac{T_N}{(T_N + F_P)}$$ **Equation 4:28**

$$Accuracy = \frac{(T_P + T_N)}{(T_P + T_N + F_P + F_N)}$$ **Equation 4:29**

• TP are true positives means number of key players which are identified correctly.

• TN are true negatives means number of normal members of network which are identified correctly.

• FP are false positives means number of normal members of network which are wrongly identified as key players.

• FN are false negatives means number of key players which are wrongly identified as normal members of network.

The modeling of classifiers has been done using randomly selected 90% of data as training and remaining 10% data as testing. The experiments are repeated 10 times and their average results are given. Table 4.6 shows the summary of results.

S.NO	CLASSIFIERS	ACCURACY
1.	Naive Bayes	92.95 %
2.	Decision Tree:	
	ID3	85.97 %
	CHAID	49.07 %
	Decision Stump	91.30 %
3.	K Nearest Neighbour :	
	K: 15	79.8 %
	K: 20	80.47 %
	K: 22	81.29 %
	K: 25	81.64 %
	K:30	82.02 %
	K: 35	82.15 %
	K:50	84.22 %
	K: 65	85.08 %
	K: 75	85.94 %
	K: 85	86.38 %

	K: 95	**86.66 %**
4.	**VOTE BASED:** **1. Naive** **2. KNN (95)** **3. Bayesian** **4. ID3**	**93.36%**

Table 4.6: Comparison of Accuracies achieved by proposed and individual classifiers

Chapter 5 Conclusions and Future work

This chapter concludes the work done and contributions made in the area of counter terrorism using latest information and communication technologies. Directions about future work have also been included for interest to carry forward this work.

5.1 Conclusions

Area of counter terrorism using latest information and communication technologies has become a very active area of research because of saving numerous human lives from dreadful terror. As mentioned in detail in the previous chapters, numerous researchers have applied various techniques covering different aspects of counter terrorism. Very bright among them are the use of social network analysis to analyze terrorist organizations structures in order to device their destabilization strategies, use of several data mining and machine learning techniques in order to identify patterns of interest which may be used by law enforcement agencies to treat with such organizations and so on.

In this dissertation, the contributions have been made in this wide area utilizing some very suitable techniques from the fields of social network analysis, outlier detection and data classification.

A new SNA measure "Relative Degree" has been proposed which can be used to identify the group leaders of terrorist networks. In terrorist networks, group leader are the individuals who keep themselves covert due to secrecy by keeping themselves least connected to others. Because of their least connectivity, they have very low values of traditionally used SNA measures like well known Degree, Closeness, Betweenness and Eigenvector Centrality. Also these individuals are directly and strongly connected with the individuals having highest values for traditional SNA measure. Relative Degree has been designed keeping these both considerations. Using Relative Degree, individuals of a network who have lowest values of traditional measures but are strongly and directly connected to the highest profile individuals are detected out. The proposed idea has been validated on various public and self detected real covert networks. As part of the same framework, a component to predict a potential terrorist event has been proposed using the area of outlier detection. Outlier detection has been chosen for this purpose because of its previous applications in similar areas like fraud detection, intrusion detection etc. This novel idea

of generating an alarm on outlying behavior of a system has been tested on a real self created dataset of cyber attackers. The maximum achieved by the proposed model was found to be 94%.

Another novel contribution was made in the area of key player detection in social network analysis of terrorist networks. As present in literature, all the traditional measures of SNA used to identify key players consider different aspects of the network, so resulting importance score using different measures can vary. To utilize benefits of all proven well known measure, a new Model for Key Player Detection using Hybrid Classifier has been proposed. A hybrid classifier, which is an ensemble of different well known classifiers including K Nearest Neighbor, Gaussian Mixture Model and Support Vector Machine was designed and proposed for this purpose. The proposed hybrid classifier was implemented and tested on three different datasets including the very famous 9 11 data set, Noor din Muhammad Top data set and the same mentioned real data set which has been created by our IT department. The detailed quantitative and comparative analysis of proposed model was performed. The performance of proposed system was measured using sensitivity (sen), specificity(spec), accuracy (acc) and area under receiver operating characteristics (ROC) curves (AUC) as figures of merit. Sensitivity is true positive rate and specificity is true negative rate. The highest accuracy achieved by the proposed hybrid classifier was found to be 95.03% which was found better than achieved by individual classifiers and other methods of combing classifiers like ada boosting and bagging.

Another important work as part of this dissertation has been made in the form of proposal of a new Terrorist Group Prediction model. The model was constructed from different classifiers including decision tree, naïve bayse and K nearest neighbor. The proposed model has a vote based ensemble of these mentioned classifiers in its core. The proposed model was implemented and experimented on publicly available global terrorism database (GTD) which has been constructed by university of Maryland. The dataset contains data about terrorism incidents since 1970 till now. The dataset and the proposed model was used to detect the terrorist organization behind any terrorist event using features like Month, City, Country, Attack_Type, Group, Target and Weapon. The results was measure in terms of sensitivity (sen), specificity(spec), accuracy (acc) and area under receiver operating characteristics (ROC) curves (AUC). The proposed vote based ensemble model achieved an accuracy of 93.36% which was higher than all

other individual classifiers. Following section covers the highlights of major contributions mode in the dissertation.

5.2 Contributions

The research contributions made in this thesis are listed as under:

- A new social network analysis measure "**Relative Degree**" to detect group leaders in terrorist networks
- Application of Outlier Detection for event prediction suspicious networks
- Construction of real dataset and used for experimentation of outlier detection for event prediction
- A novel **hybrid classifier** based key player detection ensemble model
- A novel **hybrid classifier** for prediction of responsible terrorist group in a terrorist incident.

5.3 Future Work

The work done in the dissertation can be continued adding contributions from the following fields.

5.3.1 Network Destabilization Using Link Analysis

As mentioned earlier in the previous sections, a terrorist network is a specialized type of social network because focus is on secrecy and efficiency. These networks are covert in nature and are intentionally structured to ensure efficient communication within group without being detected. Literature presents many different methods of finding the key nodes from any social network applying social network analysis which can be eliminated to destabilize a terrorist network. In the past almost all the research in the field of terrorist network analysis focuses on analysis of nodes. Links are often ignored while analyzing a network which is against the fact that links between the nodes provide at least as much relevant information about the network as the nodes themselves. If both nodes and links are given importance while analyzing a social network, especially a terrorist network, more useful information can be derived as far as destabilization strategy is concerned.

Other than the basic parameters of measuring social networks like size, density, nodal density and so on, one has to have some additional parameters because of the different nature and different point of view of analysis for terrorist networks. The two basic parameters presented in literature are secrecy and efficiency. The significance of secrecy to terrorist network is obvious, they want to protect themselves from being caught and efficiency is also important in order to achieve their objectives.

Lindelauf et al proposes a measure of secrecy which is defined by two parameters, first is the probability of exposure and the other is the link detection probability. The probability of exposure is related with individual nodes which depend upon their location in the network defined as the probability of a member of the network to be detected as a terrorist. While probability of link detection is the chance of exposure of a part of the network if a member is being detected.

According to this mentioned definition, the safest structure of a terrorist network would be a path graph where the nodes will only know their two adjacent neighbors. But if we analyze this structure from an information exchange perspective, there will be poor response as information has to travel a long distance from one end to the other of the network. This slow communication, i.e. slow information travel will lower the efficiency of the network.

Efficiency is a measure to test in a quantified way, how efficiently the nodes of a network can exchange information.

The above mentioned features can be taken as standards of validating destabilization strategy. Also all previous measures have focused on node properties. Just one contribution has been found in analyzing links instead of nodes. This can be extended which can improve the results in terms of efficiency.

5.3.2 Real time systems for event prediction

The datasets used in the experimentation of event prediction were offline datasets because of certain constraints. The software prototypes developed and used as proof of concepts were just prototypes but not working professional software. What can be done as extension of this work is to include the concept in online real time systems and develop professional online monitoring systems that can generate event prediction alarms in real time.

References

[1] "http://www.johnstonsarchive.net/terrorism/globalterrorism1.html," [Online]. Available: http://www.johnstonsarchive.net/terrorism/globalterrorism1.html. [Accessed Aug 2013].

[2] "http://www.state.gov/j/ct/rls/other/des/123085.htm," [Online]. Available: http://www.state.gov/j/ct/rls/other/des/123085.htm. [Accessed Aug 2013].

[3] S. Wasserman and K. Faust, "Social Network Analysis in the Social and Behavioral Sciences," *Social Network Analysis: Methods and Applications,* p. 1–27, 1994.

[4] V. Krebs, "Connecting the Dots -- Tracking Two Identified Terrorists," 2002.

[5] B. a. Faulkner, 1993.

[6] V. E. Krebs, "Mapping Networks of Terrorist Cells," *INSNA,* vol. 24, no. 3, pp. 43-52, 2002.

[7] A. H. Dekker, "Centrality In Social Networks: Theoretical And Simulation Approaches," *SimTecT,* 2008.

[8] "http://www.orgnet.com/sna.html," [Online]. Available: http://www.orgnet.com/sna.html. [Accessed April 2013].

[9] C. Kadushin, "Who benefits from network analysis: Ethics of social network research," *Social Netw,* vol. 27, no. 2, p. 139–153, 2005.

[10] J. S. a. S. W. P. J. Carrington, "Models and Methods in Social Network Analysis," *Cambridge, U.K. :Cambridge Univ. Press,,* 2005.

[11] L. C. Freeman, " The Study of Social Networks," [Online]. Available: http://www.insna.org/INSNA/na_inf.html. [Accessed April 2013].

[12] Scott., 1992.

[13] N. E. Friedkin, "Theoretical foundations for centrality measures," *American Journal of Sociology,* 1991.

[14] S. P. Borgatti, "Identifying sets of key players in a social network," *Comput Math Organiz Theor,* pp. 21-34, 2006.

[15] L. C. Freeman, "Centrality in Social Networks, Conceptual Clarification," *Elsevier,* 1978.

[16] T. R. Zaman, Information Extraction with Network Centralities: Finding Rumor Sources, Measuring Influence, and Learning Community Structure, Massachusetts Institute of Technology, 2011.

[17] J. Niemincn, "On the centrality in a directed graph," *Social Science Research ,* pp. 371-378, 1973.

[18] M. B. R. P.-S. a. A. V. A. Barrat, "The architecture of complex weighted networks," *National Academy of Sciences,* vol. 101, no. 11, p. 3747–3752, 2004.

[19] F. A. ,. S. Tore Opsahl, "Node Centrality in Weighted Networks: Generalizing Degree and Shortest Paths," *Social Networks,* 2010.

[20] G. Sabidussi, "The centrality index of a graph," *Psychometrika,* p. 581–603, 1966.

[21] L. Freeman, "Centrality in social networks," *Social Networks,* p. 215–239, 1979.

[22] L. C. Freeman, "A Set of Measures of Centrality Based on Betweenness," *Scoiometry,* vol. 40, no. 1, pp. 35-41, 1979.

[23] P. B. a. P. Lloyd, "Eigenvector-like measures of centrality for asymmetric relations," *Social Networks,* vol. 23, pp. 191-201, 2001.

[24] A. Dekker, "Conceptual Distance in Social Network Analysis," *Journal of Social Structure,* vol. 6, no. 3, 2005.

[25] P. &. H. F. Hage, "Eccentricity and centrality in networks," *Social Networks,* vol. 17, p. 57–63, 1995.

[26] L. B. S. &. W. D. Freeman, "Centrality in valued graphs: A measure of betweenness based on network flow," *Social Networks,* vol. 13, p. 141–154, 1991.

[27] G. Barbian, "Trust Centrality in Online Social Networks," in *European Intelligence and Security Informatics Conference*, 2011.

[28] N. M. a. H. L. Larsen, "Investigative Data Mining Toolkit: A Software Prototype for Visualizing, Analyzing and Destabilizing Terrorist Networks," in *Visualising Network Information (pp. 14-1 – 14-24). Meeting Proceedings. RTO-MP-IST-063, Paper 14. Neuilly-sur-Seine*, France, 2006.

[29] J. W. L. R. a. T. Guo, "A new measure of node importance in complex networks with tunable parameters," in *IEEE*, 2008.

[30] J. S. a. J. Adibi, "Discovering Important Nodes through Graph Entropy, The Case of Enron Email

Database," in *Illinois ACM*, Chicago, 2005.

[31] J. Korner, "Bounds and information theory," *SIAM Journal on Algorithms and Discrete Mathematics,* vol. 7, pp. 560-570, 1986.

[32] D. Gomez, "Centrality and power in social networks: a game theoretic approach," *Mathematical Social Sciences,* vol. 46, pp. 27-54, 2003.

[33] R. L. a. I. Blankers, "Key player identification : a note on weighted," in *International Conference on Advances in Social Networks Analysis and Mining*, 2010.

[34] K. M. Carley, "Destabilization of Covert Networks," *Computational & Mathematical Organization Theory,* vol. 12, no. 1, pp. 51-66, 2006.

[35] M. O. Jackson, "The Stability and Efficiency of Economic and Social Networks," *Networks and Groups,* pp. 99-140, 2001.

[36] N. R. S. Y.Narahari, "Determining the Top-k Nodes in Social Networks using the Shapley Value," in *Int.Conf.on Autonomous Agents and Multiagent Systems* , Portugal, 2008.

[37] S. Karthika, "Identifying Key Players in a Covert Network using," in *IEEE International Conference on Recent Trends in Information Technology*, 2012.

[38] J. Farley, "Breaking Al Qaeda Cells: A Mathematical Analysis of Counterterrorism Operations (A Guide for Risk Assessment and Decision Making)," *Studies in Conflict and Terrorism,* vol. 26, pp. 399-411, 2003.

[39] D. B. Skillicorn, "Social Network Analysis via Matrix Decompositions : al Qaeda," School of Computing Queen's University, 2004.

[40] J. R. N. K. Kathleen M. Carley, "Destabilizing Terrorist Networks," in *NAACSOS*, Pittsburgh, 2003.

[41] "http://www.elearningpost.com/articles/archives/qa_with_professor_karen_stephenson/," [Online]. Available: http://www.elearningpost.com/articles/archives/qa_with_professor_karen_stephenson/. [Accessed May 2013].

[42] "http://www.stratfor.com/weekly/terrorism-and-exceptional-individual," [Online]. Available: http://www.stratfor.com/weekly/terrorism-and-exceptional-individual. [Accessed 2013].

[43] B. C. Price, "Leadership Decapitation and the End of Terrorist Groups," *Belfer Center Programs or Projects: Quarterly Journal: International Security,* 2012.

[44] "Anti terrorism Agency – Kurdistan Iraq. The Geo security of Mosel.," PUK Media, 2010.

[45] "Bin Laden's long reach", published in "UNDERSTANDING THE CONFLICT TERRORISM," [Online]. Available: http://seattletimes.com/news/nation-world/crisis/terrorism/binladen_18.html. [Accessed 2013].

[46] M. A. Smith, "Analyzing (Social Media) Networks with NodeXL," in *C&T'09*, Pennsylvania, 2009.

[47] "http://nodexl.codeplex.com/," [Online]. Available: http://nodexl.codeplex.com/. [Accessed 2013].

[48] A. Berzinji, "Detecting Key Players in Terrorist Networks," Independent thesis Advanced level degree of Master, Uppsala University.

[49] Y. Musharbash, "Saif al-Adel Back in Waziristan: A Top Terrorist Returns to Al-Qaida Fold," *SPIEGEL ONLINE,* 2010.

[50] P. a. C. S. a. A. B. Sun, "Mining for Outliers in Sequential Databases," in *SIAM International Conference on Data Mining*, 2006.

[51] A. B. a. V. K. VARUNCHANDOLA, "Outlier Detection: A Survey".

[52] "http://www.graphpad.com/guides/prism/5/userguide/prism5help.html?reg_graphing_outliers.htm," [Online]. Available: http://www.graphpad.com/guides/prism/5/userguide/prism5help.html?reg_graphing_outliers.htm. [Accessed 2013].

[53] T. Q. P. R. A. S. James M. Whitacre, "Use of statistical outlier detection method in adaptive evolutionary algorithms," in *8th annual conference on Genetic and evolutionary computation*, 2006.

[54] A. M. L. Peter J. Rousseeuw, "Robust Regression and Outlier Detection," *Wiley, John & Sons, Incorporated,* 2008.

[55] C. Preisach, H. Burkhardt, L. Schmidt-Thieme and R. Decker, "Data Analysis, Machine Learning and Applications," *Springer,* 2008.

[56] L. B. S. F. J.A.S. Almeida, "Improving hierarchical cluster analysis: A new method with outlier detection and automatic clustering,," *Chemometrics and intelligent Laboratory System,* vol. 87, 2008.

[57] 1. A.-. Z. I. S. A. A.-D. A.M. Masoud, "Fast Algorithms for Outlier Detection," *Journal of*

Computer Science, vol. 4, pp. 129-132, 2008.

[58] J. Quinlan., "C4.5: Programs for Machine Learning," in *The Morgan Kaufmann Series in Machine Learning*, San Mateo, Morgan Kaufmann Publishers, 1993.

[59] J. L. B. Zadrozny, "Outlier detection by active learning," in *ACM SIGKDD*, 2006.

[60] D. H. a. C. S. I. Steinwart, "A classification framework for anomaly detection.," *Journal of Machine Learning Research,* vol. 6, p. 211–232, 2005.

[61] A. L. a. V. Kumar, "Feature bagging for outlier detection," in *ACM SIGKDD*, 2005.

[62] C. Preisach, H. Burkhardt, L. Schmidt-Thieme and R. Decker, "Data Analysis, Machine Learning and Applications," *Springer,* 2008.

[63] Y. Q. J. Liang, "Information granules and entropy theory in information systems," *Science in China Series F: Information Sciences,* vol. 51, pp. 1427-1444, 2008.

[64] V. Kreb, "Uncloaking Terrorist Network," *First Monday,* vol. 7, no. 4, 2002.

[65] P.-N. S. M. a. K. Tan, Introduction to Data Mining, Addison-Wesley, 2005.

[66] S. C. V. a. K. Boriah, "Similarity measures for categorical data: A comparative evaluation," in *SIAM International Conference on Data Mining*, 2008.

[67] V. E. E. E. L. S. G. a. K. Chandola, "Data mining for cyber security. In Data Warehousing and Data Mining Techniques for Computer Security," *Springer,* 2006.

[68] E. M. K. a. R. T. Ng., "Finding intensional knowledge of distance based outliers," in *VLDB*, San Francisco, 1999.

[69] S. Ramaswamy, "Efficient algorithms for mining outliers from large data sets," in *ACM SIGMOD*, NewYork, 2000.

[70] F. A. a. C. Pizzuti, "Fast outlier detection in high dimensional spaces," in *Principles of Data Mining and Knowledge Discovery*, London, 2002.

[71] G. H. Orair, "Distance-Based Outlier Detection Consolidation and Renewed Bearing," in *36th International Conference on Very Large Data Bases*, Singapore, 2010.

[72] H.-P. K. ,. R. T. N. ,. J. S. Markus Breunig, "LOF: Identifying Density-Based Local Outliers," in *SIGMOD*, 2000.

[73] "http://rapidminer.com/," [Online]. Available: http://rapidminer.com/. [Accessed 2014].

[74] "http://www.inf.unibz.it/dis/teaching/DWDM09/slides2012/lesson9-Classification1.pdf," [Online]. Available: http://www.inf.unibz.it/dis/teaching/DWDM09/slides2012/lesson9-Classification1.pdf. [Accessed 2014].

[75] "www.inf.unibz.it/dis/teaching/DWDM/.../lesson9-Classification1.pdf:," [Online]. Available: www.inf.unibz.it/dis/teaching/DWDM/.../lesson9-Classification1.pdf:.

[76] "https://blog.itu.dk/SPVC-E2010/files/2010/11/chapter6_datamining.pdf," [Online]. Available: https://blog.itu.dk/SPVC-E2010/files/2010/11/chapter6_datamining.pdf. [Accessed 2014].

[77] "http://engr.case.edu/zhang_xiang/teaching/eecs435/classification.pdf," [Online]. Available: http://engr.case.edu/zhang_xiang/teaching/eecs435/classification.pdf. [Accessed 2014].

[78] "https://blog.itu.dk/SPVC-E2010/files/2010/11/chapter6_datamining.pdf," [Online]. Available: https://blog.itu.dk/SPVC-E2010/files/2010/11/chapter6_datamining.pdf. [Accessed 2014].

[79] "http://www.intechopen.com/books/biodiversity-conservation-and-utilization-in-a-diverse-world/image-processing-for-spider-classification," [Online]. Available: http://www.intechopen.com/books/biodiversity-conservation-and-utilization-in-a-diverse-world/image-processing-for-spider-classification. [Accessed 2014].

[80] "http://www.scholarpedia.org/article/K-nearest_neighbor," [Online]. Available: http://www.scholarpedia.org/article/K-nearest_neighbor. [Accessed 2014].

[81] P. A. P. a. V. DePuy, "An Overview of Non-parametric Tests in SAS: When, Why, and How".

[82] A. R. A. a. R. A. Bradley, "Rank-Sum Tests for Dispersions," *The Annals of Mathematical Statistics.*

[83] T. H. P. Cover, "Nearest neighbor pattern classification," *IEEE Trans. Inform. Theory,* vol. 13, no. 1, pp. 21-27, 1967.

[84] S. T. a. K. Koutroumbas, Pattern Recognition, Burlington: MA: Academic, 1999.

[85] P. E. H. a. D. G. S. R. O. Duda, Pattern Classification, New York: Wiley, 2001.

[86] "Squares support vector machine.," [Online]. Available: http://www.esat.kuleuven.be/sista/lssvmlab/. [Accessed 2014].

[87] N. a. S. F. E. Roberts, "Roberts and Everton Terrorist Data: Noordin Top Terrorist Network

(Subset)," 2011.

[88] Y. S. R. E. Freund, "A Decision-Theoretic Generalization of on-Line Learning and an Application to Boosting," 1995.

[89] L. Breiman, "Bagging predictors," *Machine Learning,* vol. 24, no. 2, 1996.

[90] R. Bryll, "Attribute bagging: improving accuracy of classifier ensembles by using random feature subsets," *Pattern Recognition,* vol. 20, no. 6, 2003.

Appendix

Publications

A 1: Journal Publications:

1. **Wasi Haider Butt**, Usman Qamar, and Shoab A Khan, "Hidden Members and Key Players Detection in Covert Networks Using Multiple Heterogeneous Layers", Will be published in Journal of Industrial and Intelligent Information (JIII), Volume 2, No. 2, June 2014.

2. **Wasi Haider Butt,** Shoab A Khan, "Group Leaders Detection in Terrorist Networks using Social Network Analysis," Published in European Journal of Scientific Research, ISI Indexed, Volume 119 No 2 February, 2014
3. **Wasi Haider Butt,** M. Usman Akram, Shoab A Khan and M. Younas Javed**,** "Covert Network Analysis for Key Player Detection and Event Prediction using a Hybrid Classifier", The Scientific World Journal. ISI Impact Factor 1.730

A 1: International Conference Publications:

1. **WH Butt**, SA Khan, U Qamar, "Detecting Covert Dubious Actors Using Cross Domain", Third international conference on Innovative Computing Technology (INTECH 2013), London, UK. Available at : IEEE Explorer.

2. **Wasi Haider Butt**, Usman Qamar, and Shoab A Khan, "Hidden Members and Key Players Detection in Covert Networks Using Multiple Heterogeneous Layers", 1st International Symposium on Communication and Information Theory 2013, China.